Vol. 62
1964

Psychological Bulletin

HARRY HELSON, Editor
Kansas State University

VOLUME 62, 1964

PUBLISHED MONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 SEVENTEENTH ST. N. W.
WASHINGTON, D. C. 20036

ACKNOWLEDGMENT

To the following persons who have aided in reaching decisions concerning manuscripts submitted for publication during the past year the Editor extends his thanks and appreciation:

E. R. HILGARD
E. J. PHARES
J. A. STEGER

HARRY HELSON, Editor

CONTENTS OF VOLUME 62

VOL. 62, NO. 1 JULY 1964

Psychological Bulletin

SOURCES OF REPORTS OF VISUAL AND AUDITORY SENSATIONS IN PERCEPTUAL-ISOLATION EXPERIMENTS [1]

MARVIN ZUCKERMAN AND NATHAN COHEN

Research Laboratories, Division of Endocrinology and Reproduction, Albert Einstein Medical Center, Philadelphia, Pennsylvania

The review analyzes the experiments on perceptual isolation with special reference to the phenomena of reported visual and auditory sensations. Variables analyzed include: methods of confinement and restriction, conditions of illumination, duration of isolation, set, instructions and suggestions, reporting or verbalization instructions, sleep, subject populations, prior knowledge and expectations, intelligence and personality characteristics of Ss, stress response, and methods of obtaining reported visual and auditory sensations. The relevance of some of the findings to physiological, psychoanalytic, cognitive, and social psychological theories of perceptual isolation are discussed. Variables which seem important in the phenomena discussed are set, verbalization instructions, S's alertness, and E's methods of obtaining responses.

The unexpected reports of hallucinations in the first perceptual-isolation report (Heron, Bexton, & Hebb, 1953) fascinated clinicians, theoreticians, and experimentalists, and stimulated widespread interest in other responses to isolation. From the reports at symposia (Flaherty, 1961; Solomon, Kubzansky, Leiderman, Mendelson, Trumbull, & Wexler, 1961; West, 1962) and reviews of the area (Fiske, 1961; Kubansky, 1961), it is apparent that a decade after the original study the sources of variation in reports of hallucinatory-like phenomena are still unknown. The reasons are the lack of comparability of conditions, subjects (*S*s), and procedures used by the different experimenters (*E*s), as well as the lack of agreed-upon criteria for the response variables. The purpose of this paper will be to relate the situational, subject, and response variables in studies to the reported incidences of visual and auditory sensations. Reported auditory sensations will be less thoroughly analyzed because of the lack of data on this in many studies and for reasons which will be discussed later.

The variation in terminology in this area is symptomatic of the disagreement about the crucial variables in the complex experimental situation. The situation has been called "sensory deprivation," "decreased sensory variation," "sensory isolation," "sensory alteration," "reduced sensory input," "physical iso-

[1] This investigation was supported in whole by Public Health Service Research Grant MH 06875-01, EP 7-R01 from the National Institute of Mental Health.

This paper was written while the senior author was at Adelphi College.

The authors wish to acknowledge the generous cooperation of the investigators who, in many cases, had to spend considerable time re-examining their raw data.

lation," "perceptual deprivation," and "perceptual isolation." In general, the situations isolate *S* under conditions which prevent or reduce patterned or meaningful perceptions.

The response phenomena in which we are interested are called "hallucinations" by some and "images" by others. We prefer the operational terminology of Murphy, Myers, and Smith (1962) who call the visual phenomena "reported visual sensations." In this article we will use the abbreviation "RVS" for reported visual sensation and "RAS" for reported auditory sensation.

The RVSs may vary in structuredness and meaningfulness from flashes of light to complex integrated and animated scenes. Several systems of classification have been suggested. For the purposes of this review we grouped the RVSs into two categories based primarily on their "meaningfulness": A, Meaningless (e.g., flashes of light, spots, geometric forms)—including Vernon, McGill, and Schiffman's (1958) Types I and II; Murphy and his associates' (1962) Types 1 and 2; and Zuckerman, Albright, Marks, and Miller's (1962) Types *a*, *b*, and *c*; B, Meaningful (e.g., objects, people, scenes)—including Vernon's Type III, Murphy's Types 3 and 4, and Zuckerman's Types *d* and *e*. We believe that the term "hallucination" should only be applied to the B phenomena since the A phenomena may include simple idioretinal responses or illusions while the B phenomena seem to signify a greater cortical involvement.

The RASs were similarly classified for this review. The A category includes the various first-order interpretations of the "noise" described in Zuckerman and his associates' (1962) Category *a*. The B includes the sound of human voices, presence, or music described in Categories *b* and *c*.

It was difficult to obtain the breakdown of the RVSs and RASs in many of the published studies because the authors did not classify them, or in some cases, tell the reader how many of their *S*s reported such sensations. Certain other crucial information, such as the instructions or sets given the *S*s, was lacking in some publications. Because of these deficiencies we sent out forms to the chief investigators in this area asking for specific information about experimental conditions and frequencies of *S*s reporting the A and B types of RVSs and RASs. Unfortunately, not all of our inquiries were returned and some investigators no longer had access to their data so that some studies had to be eliminated or used with partial information.

RVSs

Only three of the studies failed to find at least one *S* giving an RVS. Two of these three studies were from the series of experiments by Vernon, Marton, and Peterson (1961). Six studies failed to find Type A RVSs and four of them were from the studies by these authors. Eleven studies failed to find Type B RVSs and 4 of these were from the aforementioned Princeton group. The six experiments reported by this group utilize a range of experimental conditions and sets so that the paucity of "hallucinators" is difficult to explain. However, these *E*s used more stringent criteria for defining hallucinations than most other *E*s which may account for the difference between their results and those of others. Using the median percentages within groups reporting classifiable RVSs we find 43% of *S*s giving Type A RVSs and 19% giving Type B RVSs.

Method of Confinement and Restriction of Motility

Four general methods of confinement have been used: (*a*) the *S* is confined on a chair, (*b*) the *S* is confined on a bed or mattress, (*c*) the *S* is confined in a tank-type respirator like those used for poliomyelitis patients, (*d*) the *S* is suspended in a tank of water.

The four groups where *S*s were confined in a chair (Cohen, Rosenbaum, Dobie, & Gottlieb, 1959; Goldberger, 1961; Silverman, Cohen, Shmavonian, & Greenberg, 1961) yielded practically no Type B RVSs. Two of these groups were run in darkness and two in diffused or subdued light. The periods of confinement were short, ranging from 40 minutes to 2 hours; but, as will be shown later, the period of confinement cannot account for the lack of Type B RVSs in these groups.

The results in experiments using bed or mattress confinement range from no Type A or B RVSs, as in Vernon, Marton, and Peterson's (1961) first and fifth groups to 57% of the *S*s producing Type B RVSs in the Goldberger and Holt (1958) study. Freedman, Grunebaum, and Greenblatt (1961) have pointed out that in five experiments where motility was restricted RVSs were obtained, while in two experiments (Ruff, Levy, & Thaler, 1961; Vernon et al., 1961) where motility was free or mildly restricted minimal or no RVSs were obtained. Ruff et al. (1961) concede that the lack of restriction on motility may account for their lack of RVSs. Freedman et al. (1961) postulate that the kinesthetic feedback from movement may interefere with activity in the visual modality areas of the brain. This hypothesis is similar to Rorschach's hypothesis of an inverse relationship between external movement and internal fantasy activity as reflected in the perception of human movement (*M*) in the Rorschach blots.

Supporting this restricted motility hypothesis is a significant correlation of .68 between *S*'s immobility and imagery in the Goldberger and Holt (1958) undergraduate group although the correlation was not significant in their actor group (Holt & Goldberger, 1961). Solomon and Mendelson (1962) failed to find differences between RVS and non-RVS *S*s on ratings of movement during the experiments. Bexton, Heron, and Scott (1954) observed that while complex mental activity inhibited RVSs, physical exercise or talking to *S* did not. Doane (1955) put goggles on four of his *S*s but did not restrict their audition or motility. Two of these four *S*s had RVSs and one of them had an RVS while being taken for a walk! In his total isolation group, Doane found that *S*s reported increases in RVSs during periods of activity or discomfort. Courtney, Davis, and Solomon (1961) varied the motility factor by requiring one group to make large body movements every few minutes while another group made small finger movements on the same schedule. Both groups were confined to a bed in diffuse light conditions and told to report any images. The same number of *S*s in each group had RVSs. Except for the Goldberger and Holt (1958) correlation these results fail to support the restricted motility hypothesis. Doane's observations even suggest the opposite hypothesis, that some movement may be conducive to RVSs.

Vernon et al. (1961) found their greatest incidence of Type A RVSs in the group that was taken out and led down a long corridor to go to the toilet. Groups who went to the toilet within the isolation room had few RVSs although some had Type B RVSs which were absent in the prior groups. The *S*s in Goldberger and Holt's (1958) first

study were taken out while Ss in their second study (Holt & Goldberger, 1961) were confined entirely to the room, but both groups had about the same proportion of Ss giving Type B RVSs. The Ss in the Bexton et al. (1954) and Doane (1955) experiments were taken out and both studies had relatively high proportions of Ss with RVSs. Arnoff, Leon, and Brownfield's (1962) Ss were taken out but no RVSs were obtained in the study.

The results do not support the restricted motility hypothesis or its converse. It is possible that S's actual position while in the bed may be more important than the amount of occasional exercise. A S lying on his back looking up may be more likely to produce RVSs than a S lying on his side or stomach simply because the former posture is more closely associated with states of restful alertness and scanning while the latter postures may be more associated with sleep and withdrawal of interest in the environment.

The tank-type respirator restricts gross body activity somewhat more than a bed although Ss may still flex their arms or legs within the respirator. Experiments by Solomon and Mendelson (1962); Davis, McCourt, and Solomon (1960); Davis, McCourt, Courtney, and Solomon (1961); and Zuckerman et al. (1962) have used this method of confinement. All of these studies have produced some Ss with Type B RVSs, the proportions varying from 5% to 42%. The lower figure was in the experiment by Davis et al. (1961) where husbands and wives were confined in adjoining respirators. Too much "togetherness" appeared to dampen the RVSs in this group although pairs of strangers in the same situation produced a 30% proportion for the B type RVS. The groups with social as well as perceptual isolation gave figures 18, 27, 20, and 42%. The two studies which suspended Ss in a tank of water following the Lilly model (Bliss & Clark 1962, also reported by Cambareri 1959, Shurley 1962) also obtained moderate proportions of Ss with Type B RVSs (25%, 37%).

Although the more severe methods of confinement seem to produce RVSs more consistently the results cannot be definitely attributed to kinesthetic feedback. Perhaps some other aspect of the more confining situations such as the consistent posture of S, his total helplessness, or the more unusual nature of these methods of confinement play some role.

Conditions of Illumination

Three types of visual restriction have been used: (*a*) total darkness; (*b*) diffuse light, achieved by semitranslucent eyecups or goggles which admit light but interfere with patterned form perception; (*c*) subdued light with a homogeneous field, achieved by surrounding S with a screen or dome-type *Ganzfeld*. In the latter condition, typical of the tank-type respirator studies, the S can still see his immediate surroundings and parts of his body.

A number of investigators have varied this aspect of the situation within their own experiments. Heron (1961) put opaque goggles on three Ss who had been wearing translucent goggles for several days and "hallucinating persistently." When first placed in the dark the RVSs appeared more vivid but within 2 hours disappeared or were greatly diminished. When the translucent goggles were put on again the RVSs reappeared. Two Ss were run from the beginning with opaque goggles and one of them had RVSs. When translucent goggles were placed on both Ss near the end of the experiment both reported "vivid" RVSs. Freedman and Greenblatt (1960) ran 10 Ss under darkness and 10 under

diffuse light conditions. The proportions of *S*s reporting Types A and B RVSs were almost identical in the two groups. Four *S*s were run under both of the two conditions: two had RVSs in both conditions, one in neither condition, and one in darkness only. Cohen et al. (1959) ran four *S*s with blacked-out goggles and six *S*s with frosted goggles admitting diffuse light. Type B RVSs were absent in both groups, and Type A RVSs were slightly more frequent in the diffuse light condition. Ruff et al. (1961) found only two Type B RVSs in 63 *S*s run under varying conditions; one of these occurred in a diffuse light, the other in a darkness condition. Zubek and his associates (Zubek, Aftanas, Hasek, Sansom, Schludermann, Wilgosh, & Winocur, 1962; Zubek, Pushkar, Sansom, & Gowing, 1961) ran one group (sensory deprivation) in darkness and silence and another group (perceptual isolation) with diffuse light and white noise stimulation. Fewer *S*s reporting RVSs were found in the latter group.

Vernon et al. (1961) reported on three groups run in darkness and silence conditions and three groups in diffuse light conditions. At the time when the first three groups were run (Vernon et al., 1958) the authors felt that the greater number of *S*s with RVSs in the second group was due to a failure in the blackout conditions while *S*s were being taken to the bathroom blindfolded. They concluded that nonpatterned visual stimulation might be necessary for producing RVSs. To test this they changed the conditions in the fourth and fifth groups, introducing diffuse light stimulation in the fourth group, adding unpatterned sound stimulation to this in the fifth group, and adding varying unpatterned light stimulation (flashing panel displays) in the sixth group. Type A RVSs were not found in any of these three groups but a few Type B RVS *S*s were found in the fifth and sixth groups. It is just possible that constant visual stimulation may have differential effects on the Types A and B RVSs, decreasing the former and increasing the latter.

Davis et al. (1960) tested the hypothesis that variable nonpatterned stimulation might increase RVSs. They ran a group similar to Vernon and his associates' (1961) Group VI, in which lights were flashed on a random schedule and Rorschach cards were flashed on the wall for brief intervals. Comparing their results with other respirator confinement studies it is obvious that they did not increase RVSs above the usual percentages reported in these studies. Vosburg, Fraser, and Guehl (1960) ran the same *S*s in three conditions in successive hours during the same experiment: darkness and silence, diffuse light and silence, "noise" auditory stimulation and darkness. They found no difference between the sound and visual stimulation hours in the proportion of auditory or visual sensations reported, and only slight differences between the "no stimulation" and "stimulation" hours. The authors' failure to classify their RVSs may have obscured differences in the specific types of RVSs as a function of stimulation.

Is there any difference in the results in groups run in diffuse light conditions, where the visual field is entirely limited to the homogeneous stimulation, and the subdued light conditions where *S* can see parts of the apparatus and his own body in the periphery of the field? Goldberger (1961) found fewer RVSs in a *Ganzfeld* situation than in Goldberger and Holt's (1958) diffuse light perceptual-isolation group. However, the former group was in a different confinement condition (sitting) and was exposed to the visual restriction for a much shorter period of time. The subdued light conditions in the Solomon and Mendelson (1962), Davis et al. (1960,

1961) respirator confinement experiments have produced RVS incidences comparable to the typical results of the diffuse light and darkness condition groups. Zuckerman and his associates' (1962) respirator confinement group was run in darkness and it produced only a slightly higher incidence of Type B RVSs although the difference in the Type A RVSs was considerably higher. Again this points to the possibility of darkness affecting the Type A RVSs more than the Type B RVSs. Pollard, Uhr, and Jackson (1963a, 1963b) ran one group where the total field was restricted by goggles admitting diffuse light and another group restricted by a dome (subdued light). A slightly higher incidence of both types of RVSs was found in the dome group, but the differences are certainly not significant.

It is sometimes assumed that the lack of visual restriction will result in no RVSs. This assumption about the base line for RVSs may be fallacious. It overlooks the role of immobilization and spontaneously occurring RVSs taking place in everyday life. Jackson and Pollard (1963) found that 23% of a group of *S*s responding to a questionnaire reported unusual visual experiences in their history. We do not know how many of these may have occurred in conditions of partial isolation. Zuckerman, Levine, and Biase (1964) ran a control group confined to the isolation chamber without visual restriction and 23% of *S*s reported at least one Type A RVS. None reported Type B RVSs, although these were reported in groups run in darkness, and darkness with sound. Zubek et al. (1962) found that 8% of 40 *S*s in a recumbent-control group, who were not visually restricted except by their position, reported Type A RVSs. This is not significantly different from the 10% reporting Type A RVSs in their diffuse-light group, but it is significantly different than the 56% in their prior-darkness group. In a later study (Zubek, Aftanas, Kovach, Wilgosh, & Winocur, 1963) on the effects of more severe immobilization without perceptual isolation, they confined *S*s to a foam-rubber lined coffinlike box and found that 13% of their 40 *S*s reported Type B RVSs. However, four of these five RVSs occurred during the quiet of the evening when the lights were turned down low so that some degree of perceptual isolation was in effect. RVSs have been found in poliomyelitis patients confined in respirators under similar conditions (Mendelson, Solomon, & Lindemann, 1958).

Duration of Isolation

The isolation periods range from 30 minutes for the Murphy et al. (1962) control group to 2 weeks for the three heroic *S*s in the Zubek, Welch, and Saunders (1963) EEG study. The mean time of the onset of RVSs for *S*s in the Bexton et al. (1954) study was 33 hours and in the Doane (1955) study was 30.6 hours. These authors noted that complex imagery was reported only after long periods of isolation. Since these original experiments, considerable evidence has accumulated indicating that long periods of isolation are not necessary, or even sufficient, to elicit RVSs. Ruff et al. (1961) and Vernon et al. (1961) have reported minimal Type B RVSs in experiments ranging from 48 hours to 6 days of isolation. Zubek et al. (1962) have reported few Type B RVSs in experiments of 1- or 2-week durations. Arnhoff et al. (1962) report no RVSs in a 2-day experiment. On the other hand, Goldberger and Holt (1958), Holt and Goldberger (1961), Freedman and Greenblatt (1960), and Zuckerman et al. (1962) have reported 30–57% of their *S*s giving Type A RVSs within 6–8 hour periods of isolation. Jackson

and Kelly (1962) report a high incidence of RVSs in 1 hour of isolation. They attributed their results to their experimental suggestions, but the results of Zuckerman and Cohen (1964) indicate that comparable results can be obtained in 1 hour without direct suggestion. Zuckerman et al. (1962) plotted the incidence of RVSs in successive hours of a 6-hour isolation period and found that the highest incidence of reports was during the first hour and that RVSs dropped off in succeeding hours. The drop after the first hour was sharper for unstructured RVSs than for structured RVSs which fell off more gradually. These results may have been a function of the generally decreasing verbalizations of *S*s over the 6-hour period. Pollard et al. (1963b) also found fewer verbal reports during the last 3 hours of an 8-hour isolation period, and even fewer verbal reports in a second 8-hour isolation session. Decreasing verbalization over time was also found in social isolation (Walters & Henning, 1962). This decreasing verbalization phenomenon may be an adaptation and stress reduction effect as suggested by Pollard et al. (1963b), or a reduction in interest and motivation in a monotonous environment as suggested by Zuckerman et al. (1962). Perhaps the verbal responding is extinguished because of the lack of social reinforcement from others. But whatever the cause of the phenomenon it is bound to affect the spontaneous reports of RVSs in medium or long-term isolation studies.

Murphy et al. (1962) provide the most crucial study on the effect of the duration of isolation. Instead of allowing continual reporting they interrupted at 48, 72, and 96 hours, asking *S*s to give reports of any visual sensations during 30-minute reporting periods. A control group, who had not been in isolation, was put in for 30-minute periods at the corresponding times. The isolation and control *S*s did not differ in the frequency or complexity of RVSs at any of these periods. In other words, alerted *S*s lying in the dark for 30 minutes with a set to report RVSs produce as many as *S*s who have been lying in the dark for 48, 72, or 96 hours! However, on a postexperimental questionnaire, which asks *S*s to describe their experiences during the entire period of the experiment, isolation *S*s reported more frequent and more complex RVSs than the control group. These authors (Murphy et al., 1962) conclude:

> The important aspect of isolation was that it provided cubicle Ss with lengthy waking periods in the dark during which time more frequent and more complex RVSs occurred than was the case for control Ss whose waking hours took place more often in a lighted, everyday world [p. 54].

Although this explains the differences in results on concurrent and retrospective measures it does not explain why *S*s in other experiments who were exposed to long waking periods in the dark failed to report complex RVSs after emerging from isolation. Another aspect of the data in this study which will be discussed in a later section may cast more light on this puzzle.

Set, Instructions, and Suggestion

Jackson and Pollard (1962) have postulated that the suggestions of the *E*s contained in their instructions to *S*s may be a crucial variable in eliciting RVSs. The *E*s instructions may do this in two ways: (*a*) by alerting *S* to the possibility of such effects by asking him to report them; (*b*) by creating positive or negative attitudes toward the effects, resulting in oversensitization in one case and response suppression in the other.

Myers and Murphy (1962) told one group that RVSs were found only in psychiatric patients and another group

that RVSs were normal. The former group produced significantly fewer and less complex RVSs than the latter, but it is interesting that some *S*s in the former group did report RVSs despite the strong negative connotations attached to such reports. Jackson and Kelly (1962) created a positive set for RVSs by giving their *S*s an extended talk on the expected effects of a new drug (a placebo) which supposedly combined with isolation to produce hallucinations and other changes. They attached various positive values to the ability to produce RVSs. They obtained 86% Type A and 43% Type B RVSs in only 1 hour of isolation. The lack of a control group exposed to isolation without suggestion made their results equivocal with regard to the role of suggestion. Zuckerman and Cohen (1964) used a control group (I), a group given a mild suggestion that they would have RVSs and that these were normal (II), a group given the mild suggestion and a drug (placebo) suggestion (III), and a group (IV) given the extended suggestion and the drug suggestions used by Jackson and Kelly (1962). The only significant difference found was in the Type A RVSs which increased in occurrence from Group I to Group IV. The Type B RVSs showed no response to suggestion.

Pollard et al. (1963b) compared two 8-hour isolation groups given no suggestion except to report "anything unusual," with a 3-hour group given information about RVSs and told that intelligent people report them sooner. More Type A and Type B RVSs were found in the suggestion group. In a later drug experiment (Pollard et al., 1963a) these authors found practically no difference between a group given neutral instructions with a placebo and a group given suggestion with a placebo. Apparently the suggestions "take" with some groups of *S*s and not with others. It may be difficult to generalize about the effects of suggestion without considering the type of *S*s and the specific verbal set used.

It is difficult to account for the results of many of the experiments on the basis of the sets given *S*s. In the original McGill work (Bexton et al., 1954) no one expected RVSs and their occurrence was impressed on the *E*s by spontaneous, nonsolicited reports. Solomon and Mendelson's (1962) *S*s were told that the purpose of the experiment was to see what happens to normal people and were given no special report instructions, but Type B RVSs were obtained. Goldberger and Holt's (1958) *S*s were told to report "thoughts and feelings," but a high proportion also reported Type B RVSs. Arnoff et al. (1962) told their *S*s that "hallucinations" sometimes occurred as a result of isolation, but they obtained no RVSs of any type. Cohen et al. (1959) told their *S*s that they would have visual sensations and some *S*s reported Type A RVSs but none reported Type B RVSs. Zubek et al. (1961, 1962) told their *S*s that they might have "unusual experiences" and to report them. During isolation they also asked *S*s to visualize scenes and describe them. This encouragement of visual imagery could have had suggestive effects on RVSs, but they obtained few *S*s giving Type B RVSs in their two experiments. Zubek does report that *S*s experienced unusually vivid voluntary imagery. Rossi, Sturrock, and Solomon (1963) compared the vividness of voluntary imagery reported by *S*s after hypnosis, placebo suggestion, sensory deprivation, and normal conditions. Although hypnosis enhanced imagery, placebo suggestion had no effect and sensory deprivation resulted in significantly lower vividness of imagery. They conclude that imagery in sensory deprivation cannot be explained by the hypothesis that sug-

gestion enhances visual imagery in isolation.

Another approach to the problem of suggestion is to attempt to suggest sensory changes to *S* while he is in isolation. Walters and Quinn (1960) demonstrated that *S*s exposed to sensory and social deprivation had a greater response to an autokinetic stimulus than *S*s exposed to either social or sensory deprivation only. Zuckerman et al. (1962) found that they could suggest the appearance of a "light" to most of their *S*s at the end of an isolation experiment, but that the latency of reporting the nonexistent light did not correlate with the prior RVSs obtained during isolation. Bexton et al. (1954) found that their *S*s had poor control over their RVSs. Some *S*s were disturbed by them but could not stop or start them at will. There was some control over content; *S* might see some objects suggested by *E* but not always as *E* intended and with some irrelevant RVSs intruding themselves.

Continuous Reporting Instructions

The decreasing number of *S*s giving RVSs in each successive hour of isolation in the Zuckerman et al. (1962), Vosberg et al. (1960), and Pollard et al. (1963b) studies may have been related to the decrease in spontaneous verbalizations in these samples with successive hours of isolation. Since we had verbal productivity measures on *S*s in the Zuckerman et al. (1962) experiment, we correlated this measure with the RVS measure combining frequency and complexity of RVSs. The correlation was .43. Goldberger and Holt (1958) found a .87 correlation between verbal output and imagery in their male undergraduates! In their group of unemployed actors (Holt & Goldberger, 1961), the correlation was insignificant. We analyzed the data in the Pollard et al. (1963b) study and found that in their Dome group, where verbal output is given for individual *S*s, *S*s giving RVSs emitted more than three times the number of words than *S*s not giving RVSs; the difference was significant. Solomon and Mendelson (1962) reported no difference between hallucinators and nonhallucinators on verbalization.

Instructions which forbid, discourage, or limit spontaneous verbal reporting may inhibit RVSs. Vernon et al. (1961) did not permit reporting during their first three experiments and obtained no RVSs of the B type, although a large proportion of *S*s reported Type A RVSs in the second experiment. Silverman and his associates' (1961) *S*s were not told to report during the experiment and they obtained few RVSs. Zubek et al. (1961, 1962) and Arnoff et al. (1962) told *S*s they could report if they wanted to but instructions minimized reporting in these experiments and few Type B RVSs were given. Murphy et al. (1962) found significantly less complex RVSs in groups where no reporting periods were allowed during isolation than in groups where there were reporting periods during isolation; the comparisons were made on a postexperimental questionnaire.

Myers and Murphy (1962) found that prior verbalization encouraged by administering the Rorschach test before isolation had no effect on RVS complexity. Does the verbalization itself stimulate the RVSs or is it only symptomatic of the heightened interest in sensations, images, and feelings? Murphy et al. (1962) ran one isolation group where *S*s signaled RVSs by pulling a lever instead of verbally reporting them. The postexperimental scores for this group were the same as those for *S*s who reported verbally during the experiment. These results suggest that the verbal response per se is not the source of RVSs.

Sleep and Cortical Alertness

Freedman, Grunebaum, Stare, and Greenblatt (1962) feel that the RVSs in isolation are likely to occur in transitional states between sleep and waking, and may represent hypnogogic and hypnocampic imagery. If their contention is correct, then persons who doze and wake frequently during isolation should have more RVSs than *S*s who remain more alert. According to West (1962) the greater the level of arousal during reduced sensory input, the more vivid the hallucination. This theory would suggest that the more alert individual in perceptual isolation would be more prone to have RVSs than the drowsy individuals.

What is the evidence bearing on these speculations? Bexton et al. (1954) reported EEG patterns typical of "alertness" occurring during RVSs. Solomon and Mendelson (1962) reported no differences in rated sleep between RVS and non-RVS *S*s. An analysis of the Pollard et al. (1963b) data reveals no difference in time spent sleeping by RVS and non-RVS *S*s. Goldberger and Holt (1958) did not find a significant correlation between sleep and RVS imagery in their undergraduate group, but they found a −.66 correlation in their (Holt & Goldberger, 1961) actor sample. Zuckerman et al. (1962) found a −.71 correlation between sleep ratings and RVS complexity. These last two correlations indicate that *S*s who slept more frequently reported fewer and simpler RVSs and *S*s who were more alert reported more and more complex RVSs. Murphy et al. (1962) compared three groups: (*a*) an immediate RVS group where *S*s were put into a dark room and asked immediately to report RVSs, (*b*) a delayed RVS group, where *S* spent 30 minutes in darkness before being asked to report RVSs, (*c*) a wake-up group, who also waited 30 minutes in darkness before reporting, but were asked to do mental arithmetic problems for the last 7 minutes of the waiting period to eliminate drowsiness incurred during the waiting period. The delayed-RVS group more frequently reported becoming drowsy or falling asleep, and they reported less complex RVSs than the immediate or the wake-up groups.

In the studies which have found significant relationships between sleep-alertness and RVSs during isolation, alertness is found to be more conducive to RVSs and states of sleeping or drowsiness seem to be inimical to RVSs, particularly the more complex ones. This might explain why some of the long-duration experiments yield few RVSs despite the more extended opportunity to "sample" the visual field. Since the *S*s in these experiments know that they have a long time to spend in isolation, and since there are no restrictions on their position, they may simply sleep as much as they can or remain in sustained drowsy states with little attention paid to changes in their visual fields. Bexton et al. (1954) reported that *S*s tended to spend the earlier part of their experiment in sleep; later they appeared more eager for stimulation. This would explain why RVSs did not appear in the earlier hours of their experiment. In Zuckerman and his associates' (1962) groups, *S*s were noted to be most alert in the earlier hours of the experiment (*S*s were not allowed to sleep in this study) and reported the greater number of RVSs during this early period. The factor of set, used by Jackson and Kelly (1962), may increase RVSs by creating an alert, expectant attitude. If *S* is set to "scan" his visual field he is more likely to report visual sensations. At the physiological level, the relative arousal of the reticular formation may be used to explain the same phenomena.

Population Variables

The population from which *S*s for isolation experiments are drawn is some-

what limited by the fact that only certain *S*s will volunteer when informed of the experimental procedure. One wonders how specific the RVSs are to a population already motivated to experience isolation. Murphy et al. (1962) put their nonvolunteers for prolonged isolation into their control group, but it was possible to compare their RVSs in the 30-minute report periods with controls who had volunteered for isolation. There was no significant difference in the complexity of RVSs.

Within the volunteer samples the motives for volunteering differ. Wexler, Mendelson, Leiderman, and Solomon (1958) found that 5 out of 17 *S*s volunteered to "test themselves" while most of the others volunteered for the money. Four of the 5 self-testers quit the experiment before 8 hours; only 1 of the 11 mercenary *S*s could not make 8 hours. Apparently the self-testers were an unsuccessful counterphobic group. However, no relationship was found between motivation for volunteering and hallucinations (Solomon & Mendelson, 1962).

Ruff et al. (1961) speculated that his failure to obtain many RVSs may have been because his *S*s were older, experienced airmen who were accustomed to stressful activities. Murphy and his associates' (1962) soldiers had many RVSs so a military identification per se is not incompatible with RVSs. Furthermore, Ruff's assumptions that stress resistance is somehow correlated with RVSs is not borne out as will be seen in a later section.

Sophistication is also not incompatible with RVSs. Heron, Doane, and Scott (1956) used themselves as *S*s and all reported complex RVSs. Bliss and Clark (1962) used psychiatrists and psychologists and obtained a moderate proportion of RVSs. Finding RVSs in sophisticated *S*s tends to cast some doubts on the Cambareri (1959) thesis that suggestible *S*s are more prone to give RVSs. Holt and Goldberger (1961) found little difference between college students and actor samples in percentages of *S*s giving Type B RVSs although the actors gave considerably more Type A RVSs.

As with most psychological research, the undergraduate remains the favorite *S* although many *E*s also use some graduate students. No generalizations can be made about national differences since RVSs are frequently given by New York University (Goldberger & Holt, 1958) and McGill students (Bexton et al., 1954) and rarely given by Princeton (Vernon et al., 1961) and Manitoba students (Zubek et al., 1961, 1962). These differences cut across national boundaries.

Because of the theoretical analogy made between isolation and psychotic phenomena (Rosenzweig, 1959), the results of psychiatric populations are of some interest. Cohen et al. (1959) ran four normals and six patients of various diagnoses. Two of the normals and one schizophrenic gave RVSs but none of these were Type B RVSs. The RASs were more common and will be discussed later. Azima, Lemieux, and Azima (1962) reported that 7 out of 18 patients had RVSs ranging from simple to complex. They do not provide a breakdown on the types of RVSs, but the overall frequency is no higher than in most of the studies on normals. This is not surprising since spontaneous RVSs are rarer in patients than is commonly supposed. Malitz, Wilkens, and Esecover (1962) found only a 9% incidence of RVSs in case histories of 100 chronic schizophrenic patients. A case-history survey of normals might reveal a similar percentage. Cleveland, Reitman, and Bentinck (1963) found no significant difference between schizophrenic and nonschizophrenic patients in frequency or intensity of reported perceptual distortions. Schizophrenics who halluci-

nated before also admitted hallucinating in isolation; one said that the sensory deprivation made the sensations louder and clearer.

Most of the experiments have used males. Zubek et al. (1961) found suggestion of a sex difference with 10 of his 12 male *S*s giving RVSs and only 1 of his 4 females. Pollard et al. (1963b) obtained usable data on 22 males and 23 females in three conditions of isolation. Over all conditions, 73% of the males and 61% of the females gave RVSs.

*S*s *Prior Knowledge and Expectations*

Jackson and Pollard (1962) have suggested that the phenomena observed in isolation experiments are self-perpetuating because the publicity given prior studies affects the *S*s expectations and the expectancies affect their reports. Although a number of experimenters have varied experimental suggestions few have inquired about *S*s own knowledge of the area and their expectations for their own reactions before the experiments. Zuckerman and Cohen (1964) did this and found no relationship between prior knowledge and expectation, and RVSs. The *S*s specifically expecting to have hallucinations gave no more RVSs than *S*s not having such an expectation. Jackson and Pollard (1962) argue that *S* must not only expect RVSs but must be motivated to report them. Zubek et al. (1961) note that 5 of the 16 *S*s who had no RVSs were very disappointed in not having had them; surely these *S*s were motivated to report RVSs if they had experienced them.

Intelligence

Goldberger and Holt (1958) found a .59 correlation between intelligence, as measured by the Ohio State Psychological Examination, and imagery in their undergraduate sample. Murphy et al. (1962) found no difference between high and low scorers (within a restricted range of above average) on the General Technical measure of intelligence. There was a difference within the group where the postisolation inquiry was not preceded by reports during isolation. In this subgroup *S*s with high General Technical scores produced less complex retrospective RVSs. Although intelligence was not correlated with other variables in the Zuckerman et al. (1962) study, American Council on Education test scores were available, and an analysis revealed no relationship between American Council on Education measured intelligence and RVS scores.

Personality Characteristics

Holt and Goldberger (1961) have undertaken the broadest personality studies in this area. Unfortunately the generality of their results is limited by the size of the two sample studies (*N*s of 14 and 16 but sometimes dropping as low as 9 for some correlations). If the results of the second group (actors) had replicated the results of the undergraduate group, one might have more confidence in their results. However, in some cases the results were significant in the opposite direction, while in others they simply failed to confirm the correlations found in the first sample. In the undergraduate sample "imagery" in isolation correlated positively with test scores, and test and interview-based ratings, measuring "acceptance of one's passive, feminine side," "intellectual flexibility, breadth and richness," and "freedom from emotional disturbance or constriction." In the actor sample, active, masculine strivings and values correlated positively with imagery. "Intellectual flexibility" correlated with imagery in the actor sample also, but only one measure of it was significantly related in both groups. The number of *M* on the Rorschach, which correlated positively with imagery in undergraduates, failed to correlate significantly in

the actor group. Measures of emotional disturbance or neuroticism correlated negatively with imagery, as in the undergraduate sample, but the only single measure whose correlation was replicated across the samples was the Block Neurotic Under-Control scale. In retrospect it seems unfortunate that Holt and Goldberger did not use another sample similar to their first to replicate their results. If we look only at the Gestalt of the results, intellectual flexibility and freedom from emotional disturbance seem to be related to imagery in both samples while the role of masculinity-femininity and activity-passivity traits seems to depend on other characteristics of the populations.

Solomon and Mendelson (1962) found no differences between hallucinators and nonhallucinators on MMPI or EPPS scales. Zuckerman et al. (1962) found that the only MMPI or EPPS scales correlating with RVSs was the MMPI *Mf* scale. Freedman et al. (1962) found no clear relationship between ego integrity and visual imagery although they did find a significant relationship with a history of hypnogogic imagery. The failure of any of the clinical scales from the MMPI to correlate with RVSs in two studies does not support Goldberger and Holt's relationship between freedom from emotional disorder and RVSs.

Cambareri (1959) administered a battery of "suggestibility" tests to his *S*s and, on the basis of a combined rank score, divided them into suggestible and nonsuggestible groups. Out of his matrix of 45 correlations between the suggestibility tests only two were significant and most were close to zero. There was little basis for combining the tests into a composite score and therefore no basis for labeling the score "suggestibility." Whatever the score represented, 9 of the 10 high-scoring *S*s and only 2 of the 10 low-scoring *S*s reported RVSs on a postexperimental introspective report. No differences between groups were obtained on a questionnaire. It is interesting that the high scorers on the test battery found the situation more comfortable and stayed longer than low scorers.

Silverman et al. (1961) selected 11 field-oriented and 9 body-oriented *S*s, after screening 109 students with the Draw-a-Person and Rod and Frame Tests, and gave them 2 hours of isolation. Four of the field-oriented group experienced visual imagery having specific form and which they believed to come from a source outside of themselves. None of the body-oriented group reported this type of imagery. The results suggest that the field-dependent personality is more prone to RVSs, possibly because he is less able to discriminate internal and external cues in ambiguous situations.

Considering the number of correlations run between RVS scores and personality variables the results are meager and only suggestive at the most. Perhaps personality variables are important only insofar as they affect *S*s cooperation with certain of *E*s instructions, for example, to lie still, to report from time to time, not to sleep, etc. Holt and Goldberger (1961) found that imagery correlated with a group of scales that denote "being a 'good' subject: a wish to cooperate . . . persistence with the task at hand." This may have been what Cambareri was measuring in his various tasks, and would account for the fact that his high scorers remained in the tank longer than low scorers.

Stress Response and RVSs

Much of the initial excitement about the isolation RVSs was due to the assumption that they had something in common with psychotic RVSs. Actually psychotic hallucinations are usually personally symbolic and emotionally charged. While a few *S*s are frightened

by their RVSs, others are interested or amused by them, and RVSs themselves often consist of mundane objects or cartoonlike figures. Solomon and Mendelson (1962) found that in seven RVS *S*s affect was pleasant in two cases, unpleasant in three, and severe anxiety occurred in three. However, their *S*s had not been alerted to the possibility of RVSs. RVS and non-RVS groups in their experiment did not differ in somatic complaints or length of stay, two indicators of stress. They found that the non-RVS group had a larger increase in noradrenalin excretion and urine volume output than the RVS group. This might indicate that the less stressed *S*s report RVSs. Supporting this is Holt and Goldberger's (1961) findings of significant correlations between "imagery" and a cluster of variables which they call "adaptive response" to isolation. The relationship suggests that RVSs may be a self-stimulation coping mechanism rather than the projection of anxiety-laden fantasy. Zuckerman et al. (1962) found no correlation between their RVS score and any of the variables in their stress response cluster. The findings were different for RASs and these will be discussed later. Murphy et al. (1962) found that *S*s who left isolation before 48 hours reported RVSs of the same complexity as *S*s who stayed the full 96 hours. An analysis of the data in the Pollard et al. (1963b) Dome group revealed no differences between RVS and non-RVS groups in time remaining in isolation.

Methods of Reporting RVSs

There are three basic methods for obtaining RVSs: (*a*) reports during the experiment, spontaneously, or during allotted reporting periods; (*b*) a postexperimental interview, usually structured but sometimes free associative; (*c*) a postexperimental questionnaire. Some *E*s have used one, some have used two, and some have used all three methods to define the RVS phenomena. When two or more methods were used to define a phenomenon, we should ask some questions about their relationships. Zuckerman and Cohen (1964) compared RVSs in *S*s reports during the experiment and their replies to questions about RVSs in the postexperimental interview. The correlation between the two methods was .77. Although the two methods are correlated they may still yield different findings. Murphy et al. (1962) found a high order of relationship between the during and postexperiment techniques, but they found no difference between isolation and control groups on complexity of RVSs during the experiment and a significant difference on a postexperimental questionnaire. Many experimenters ignore this problem by using both the during and after reports together; that is, if *S* gives an RVS during the experiment or after the experiment it is counted. Such a procedure will certainly inflate the number of RVSs in the group over what would be obtained by using more rigorous criteria. An analysis of the data in the Zuckerman and Cohen study (1964) indicated that the use of reports appearing in both methods would give a figure of 45% RVS *S*s while the use of reports in either gives a figure of 71%. In Solomon and Mendelson's (1962) study we have a choice between 7% and 25% depending on whether we admit RVSs appearing in retrospective reports only as data. If a third method is added, as in Goldberger and Holt's (1958; Holt & Goldberger, 1961) experiments which use reports during the experiment, a postexperimental interview, and a postexperimental questionnaire, we can expect even higher proportions of *S*s giving RVSs. One is reminded of the results in the verbal conditioning experiments where the more the postexperimental in-

terview was extended, the more *S*s were found to be "aware."

Reported Auditory Sensations

The RASs have not received the attention that RVSs have. Some *E*s do not provide data on them. There are several reasons for this neglect. In rooms which are not soundproof various noises from the adjoining rooms and corridors may be reported by *S*s immediately or in the postexperimental interview. This makes it difficult for *E* to distinguish between sensation, illusion, and hallucination. Even in soundproof rooms certain reported sounds such as "water" noises are typical and may be a reaction to inner-ear noise. Low frequency sounds such as the rumble of passing trains may penetrate the soundproofing. In many experiments some kind of constant background noise such as white "noise" transmitted through earphones, the noise of the intercom system, or the hum of a motor may be used. Some *S*s begin to interpret these sounds and the interpretations may be classified as illusions rather than hallucinations. Human voices or music tend to fall closer to the hallucinatory category. The range of these Type B RASs in the groups reviewed is from 0 to 50%, with the median falling at 15%. A third of the studies report no Type B RASs. Because of the limited data available on RASs we will be somewhat more cursory in our analysis of the variables related to them.

What is the relationship between RVSs and RASs? In general, *E*s who obtained few or no RVSs also obtained few or no Type B RASs. This would suggest that there is a correlation between the two types of reported sensations. Zuckerman et al. (1962) found a .49 correlation between visual and auditory hallucination scores. This would suggest that some of the factors which influence RVSs may also influence RASs.

The RASs have been obtained in all the methods of confinement. They have been obtained under darkness and diffuse light conditions. Zuckerman, Levine, and Biase (1964) ran a group of 12 *S*s with auditory restriction, but without visual restriction, and found that 38% of *S*s gave Type B RASs in contrast to 7% in a visual and auditory restriction group. Perhaps the increased RASs in this group occurred because they were less distracted by RVSs. Pollard et al. (1963b) found fewer RAS *S*s in a subdued light (dome) condition than in a more restricting diffuse light (goggles) situation.

The role of some kind of masking noise in producing RASs was previously suggested. Zubek et al. (1961) found no Type B RASs when they used a soundproof chamber, but they obtained a few when white noise was fed to *S* through earphones (1962). Vernon et al. (1961) did not find any RASs in a group run with "noise" stimulation. Although constant auditory stimulation may play some role in RASs it is not sufficient to explain their occurrence or absence. Zuckerman, Levine, and Biase (1964) used music as sound stimulation for *S*s lying in darkness and found only slightly more Type B RASs than in the total-isolation group.

As was the case for RVSs, the duration of isolation does not seem to be a critical factor in RASs. The RASs are obtained in some 1-hour experiments and are not found in some of the 2-day to 1-week experiments.

The RASs appear more resistant to suggestion than RVSs. Vernon et al. (1961) tried to suggest to their *S*s that they would hear "music" but none of them reported hearing it. Pollard et al. (1963a, 1963b) found no more RASs with suggestion than without it. Zuckerman and Cohen (1964) found no in-

crease in RASs with increasing suggestion.

If meaningful RASs are not a function of external stimulation or suggestion they may represent projections of internal thoughts and therefore may be more closely related to personality factors than are RVSs. Since Malitz et al. (1962) found that RASs were much more frequent in schizophrenics than RVSs, we might expect that RASs would be related to deviant tendencies. Zuckerman et al. (1962) found positive but insignificant correlations between RAS scores and the Validity and Schizophrenia scales of the MMPI; however, the RAS score correlated significantly with verbalized manifest anxiety and increases in a checklist measure of anxiety during isolation. This finding suggests that RASs may reflect the situational anxiety and projective tendencies induced by isolation rather than broader personality characteristics. In the Cohen et al. (1959) study, one of four normals, two of three neurotics, and all of the three schizophrenic *S*s gave RASs. The *N*s in the group are too small to determine if the trend toward increasing RASs with increasing psychopathology is significant. Cohen et al. noted that one difference was that the normals regarded the events as "hallucinations" whereas most of the psychiatric patients regarded them as real. In Azima and his associates' (1962) psychiatric patients consisting of neurotics, depressives, and schizophrenics, only 2 of the 17, or 12% of the group gave RASs. These 2 were both neurotics. Cleveland et al. (1963) found no differences in RAS intensity or frequency between schizophrenic and nonschizophrenic psychiatric patients. In evaluating population comparisons we must again consider the question of the preisolation base line. Jackson and Pollard (1963) found that 31% of a normal sample reported "unusual auditory experiences" outside of isolation.

There was no evidence of a sex difference in RASs in the Pollard et al. (1963b) study where 64% of the men and 65% of the women *S*s gave RASs (of all types).

The comments made about the methods of obtaining reports and verbalization during isolation in relation to RVSs probably apply to RASs since the same studies, which relied mainly on postisolation indices of RASs (Arnoff et al., 1962; Silverman et al., 1961; Vernon et al., 1958; Zubek et al., 1961, 1962), or did not encourage free reporting during isolation, failed to obtain many RASs. The correlation between verbal productivity and RAS score in the Zuckerman et al. (1962) experiment was .77! In the Pollard et al. (1963b) experiment RAS *S*s spoke more words than non-RAS *S*s, but the difference on a score corrected for time in the situation and sleep time was not significant ($t=1.50$).

Theoretical Considerations and Conclusions

There have been four general approaches to the explanation of the phenomena of perceptual isolation: the physiological, the psychoanalytic, the cognitive, and the social-psychological. The physiological explanations point to specific sensory effects and to nonspecific central nervous system effects (Heron, 1961). Several investigators have cited the work of Granit (1955) which shows that spontaneous discharge of the retinal ganglion cells takes place in the dark-adapted eye as well as in the light-adapted eye. This means there is always a random "noise" factor which could furnish the sensory basis for RVSs despite the reduction or elimination of external stimulation. Doane (1955) relates the RVS isolation phenomena to RVSs in cataract cases, and to the

phantom-limb phenomenon in amputees. He suggests that the Cannon and Rosenblueth (1949) law of denervation may account for all of these cases; that is, sensory deprivation may result in a functional denervation of sensory neurons resulting in a sensitization, or lowering of thresholds, of neurons further up in the nervous system. Supporting this theory is the finding that RVSs typically progress from simple to complex (Heron, 1961; Myers & Murphy, 1962; Zuckerman & Cohen, 1964) indicating a possible progression of their site from lower to higher centers in the nervous system.

Evarts (1962) in his "neurophysiologic theory of hallucination" says that the neurophysiologic processes which produce hallucinations are similar to those which occur in dreams during sleep. Scheibel and Scheibel (1962) stress the role of the brain-stem reticular core:

> Normally periods of decreased total input to the reticular core occur only during sleep, but in this case (sensory deprivation) the individual remains nominally awake. The gradual development of n.o.b. (RVS) phenomena may represent simply a normal expression of the physiologic function of brain cells whose modulation by specific sensory and nonspecific reticular input drops below a critical level with consequent changes in cell biasing and concomitant increased "sensitivity" to what is usually "background activity" in intracerebral loops [pp. 29–30].

The former theory would suggest that RVSs occur in stages of light sleep or transitional stages while the latter theory suggests that RVSs would occur in alert, wakeful stages during perceptual isolation. The evidence from EEGs during RVSs (Heron, 1961) and from reported relationships between sleep during isolation and RVS's production (Holt & Goldberger, 1961; Murphy et al., 1962; Zuckerman et al., 1962) support the latter theory. Lindsley (1961) has also stressed the role of the "ascending reticular activating system" which may compensate for reduced sensory input by projecting its own response level upon the cortex.

The psychoanalytic theory of isolation, as delineated by Azima et al. (1962), Goldberger and Holt (1958), Goldfried (1960), and Kubie (1961), states that isolation produces a kind of regression resulting in a decrease in the efficiency of secondary process thinking (logical, problem solving, goal directed) and an increase in primary process thinking (governed by the pleasure principle, alogical). The RVSs could be explained as one aspect of primary process thinking. However, individuals who are severely threatened by primary process may develop anxiety which might interfere with RVS development. Goldberger and Holt (1958) found that a preisolation Rorschach measure of controlled primary process, correlated .45 with imagery during isolation; and controlled primary process, measured from verbalizations during isolation, correlated .88 with imagery during isolation. However, Solomon and Mendelson (1962) found that the hallucinators in their experiments daydreamed less during isolation than nonhallucinators. Perhaps this finding could be explained away by classifying daydreams as secondary process, since they are more under voluntary control than are RVSs. However, most of the imagery reported during isolation is not obviously wish-fulfilling or drive relevant. Also this theory would be hard put to explain the lack of a positive relationship between time of isolation and RVS production since ties with reality would be weakened as a function of time in isolation. The theory might explain the positive relationship between verbalization and RVSs since individuals who free-associate more easily can also release primary process more easily. In general, Goldberger and Holt's formulations are based on correlations rather than actual meas-

ures of changes in thinking during isolation. Research is needed to test the hypothesis that isolation tends to increase one type of thinking as opposed to another. The progressive decrease in verbalization during isolation would suggest that all types of associative thinking are reduced.

A third class of theory is the cognitive theory (Bruner, 1961; Freedman, 1961) which suggests that RVSs may be due to the organism's attempt to maintain "ordered relationships" in its perceptual environment. The RVSs may be the result of the organism's attempt to fit the random noise sensations from the retina, or the inner ear, into previously acquired cognitive schemata. We might deduce from this theory that the content of RVS or RASs would be dominated by the most familiar objects or sounds experienced by *S*s. Research on this theory might attempt to relate the content of RVSs of individual *S*s to their particular schemata.

The fourth class of theory is the social-psychological theory which is excellently developed by Jackson and Pollard (1962). These authors have questioned the prior theoretical interpretations of RVSs and other isolation phenomena on the basis of the fact that some *E*s have failed to obtain them, that manipulation of *S*'s set by suggestion may affect them, and that there are fewer of them during a second exposure to isolation. They advocate an explanation based on: *S*s knowledge of "expected" reactions, his motivation to experience and report, and the use of continuous reporting instructions which have a self-suggestive effect. In our review of this area we have seen that negative or positive values attached to RVSs by *E*s have affected their occurrence in some experiments but not in others. Direct suggestion to *S* that he will experience RVSs, or his expectation or desire to experience RVSs, seem to have little effect. Set probably influences RVSs but the exact role of set has not yet been clearly defined. Jackson and Pollard have used a shotgun type of suggestion, including placebos, and their failure to obtain results in their last experiment (Pollard et al., 1963a) raises questions about the role of specific instructions, in particular contexts. Furthermore, certain types of response indicators may be more sensitive than others, and certain types of RVSs may be more responsive to suggestion than others. The work so far would suggest that set alone cannot explain the RVSs, and certainly not the RASs, in isolation. Continuous reporting instructions and *S*'s responses to them certainly seem to be of crucial importance. The exact meaning of this relationship is open to question.

Most of the theoretical positions discussed above are not incompatible with each other, but could easily incorporate one another. As yet, the theoretical issues are not drawn so sharply that we could pose one theory against another. Hopefully, the next decade of work on perceptual isolation will witness more precise definition of the theories, and experiments more precisely aimed at the theoretical issues.

REFERENCES

Arnhoff, F. N., Leon, H. V., & Brownfield, C. A. Sensory deprivation: The effects on human learning. *Science,* 1962, **138**, 899–900.

Azima, H., Lemieux, M., & Azima, Fern. Isolement sensoriel, étude psychopathologique et psychanalytique de la régression et du schema corporel. *Evolut. psychiat.,* 1962, **2**, 259–282.

Bexton, W. H., Heron, W., & Scott, T. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.,* 1954, **8**, 70–76.

Bliss, E. L., & Clark, L. D. Visual hallucination. In L. J. West (Ed.), *Hallucinations.* New York: Grune & Stratton, 1962. Pp. 92–107.

Bruner, J. S. The cognitive consequences of early sensory deprivation. In P. Solomon et al. (Eds.), *Sensory deprivation.* Cam-

bridge: Harvard Univer. Press, 1961. Pp. 195–207.

Cambareri, J. D. The effects of sensory deprivation on suggestible and nonsuggestible psychology graduate students. *Dissert. Abstr.*, 1959, **19**, 1813.

Cannon, W. B., & Rosenblueth, A. *The supersensitivity of denervated structures.* New York: Macmillan, 1949.

Cleveland, S. E., Reitman, E. E., & Bentinck, C. Therapeutic effectiveness of sensory deprivation. *Arch. gen. Psychiat.*, 1963, **8**, 455–460.

Cohen, B. D., Rosenbaum, G., Dobie, S. I., & Gottlieb, J. S. Sensory isolation: Hallucinogenic effects of a brief procedure. *J. nerv. ment. Dis.*, 1959, **129**, 486–491.

Courtney, J., Davis, J. M., & Solomon, P. Sensory deprivation: The role of movement. *Percept. mot. Skills*, 1961, **13**, 191–199.

Davis, J. M., McCourt., W. F., Courtney, J., & Solomon, P. Sensory deprivation, the role of social isolation. *Arch. gen. Psychiat.*, 1961, **5**, 84–90.

Davis, J. M., McCourt, W. F., & Solomon, P. The effect of visual stimulation on hallucinations and other mental experiences during sensory deprivation. *Amer. J. Psychiat.*, 1960, **116**, 889–892.

Doane, B. K. Changes in visual function following perceptual isolation. Unpublished doctoral dissertation, McGill University, 1955.

Evarts, E. V. A neurophysiologic theory of hallucinations. In L. J. West (Ed.), *Hallucinations.* New York: Grune & Stratton, 1962. Pp. 1–14.

Fiske, D. W. Effects of monotonous and restricted stimulation. In D. W. Fiske & S. Maddi (Eds.), *Functions of varied experience.* Homewood, Ill.: Dorsey Press, 1961. Pp. 106–144.

Flaherty, B. E. (Ed.) *Psychophysiological aspects of space flight.* New York: Columbia Univer. Press, 1961.

Freedman, S. J. Perceptual changes in sensory deprivation: Suggestions for a conative theory. *J. nerv. ment. Dis.*, 1961, **132**, 17–21.

Freedman, S. J., & Greenblatt, M. Studies in human isolation. *USAF med. J.*, 1960, **11**, 1479–1497.

Freedman, S. J., Grunebaum, H. U., & Greenblatt, M. Perceptual and cognitive changes in sensory deprivation. In P. Solomon et al. (Eds.), *Sensory deprivation.* Cambridge: Harvard Univer. Press, 1961. Pp. 58–71.

Freedman, S. J., Grunebaum, H. U., Stare, F. A., & Greenblatt, M. Imagery in sensory deprivation. In L. J. West (Ed.), *Hallucinations.* New York: Grune & Stratton, 1962. Pp. 108–117.

Goldberger, L. Homogeneous visual stimulation (*Ganzfeld*) and imagery. *Percept. mot. Skills*, 1961, **12**, 91–93.

Goldberger, L., & Holt, R. R. Experimental interference with reality contact (perceptual isolation) method and group results. *J. nerv. ment. Dis.*, 1958, **127**, 99-112.

Goldfried, M. R. A psychoanalytic interpretation of sensory deprivation. *Psychol. Rec.*, 1960, **10**, 211–214.

Granit, R. *Receptors and sensory perception.* New Haven: Yale Univer. Press, 1955.

Heron, W. Cognitive and physiological effects of perceptual isolation. In P. Solomon et al. (Eds.), *Sensory deprivation.* Cambridge: Harvard Univer. Press, 1961. Pp. 6–33.

Heron, W., Bexton, W. H., & Hebb, D. O. Cognitive effects of a decreased variation to the sensory environment. *Amer. Psychologist*, 1953, **8**, 366. (Abstract)

Heron, W., Doane, B. K., & Scott, T. H. Visual disturbance after prolonged perceptual isolation. *Canad. J. Psychol.*, 1956, **10**, 13–16.

Holt, R. R., & Goldberger, L. Assessment of individual resistance to sensory alteration. In B. E. Flaherty (Ed.), *Psychophysiological aspects of space flight.* New York: Columbia Univer. Press, 1961. Pp. 248–262.

Jackson, C. W., Jr., & Kelly, E. L. Influence of suggestion and subjects' prior knowledge in research on sensory deprivation. *Science*, 1962, **132**, 211–212.

Jackson, C. W., Jr., & Pollard, J. C. Sensory deprivation and suggestion: A theoretical approach. *Behav. Sci.*, 1962, **7**, 332–342.

Jackson, C. W., Jr., & Pollard, J. C. A preliminary analysis of additional nondeprivational variables which influence research on sensory deprivation. Paper read at Midwestern Psychological Association, Chicago, 1963.

Kubie, L. Theoretical aspects of sensory deprivation. In P. Solomon et al. (Eds.), *Sensory deprivation.* Cambridge: Harvard Univer. Press, 1961. Pp. 208–220.

Kubzansky, P. E. The effects of reduced environmental stimulation on human behavior: A review. In A. D. Biderman & H. Zimmer (Eds.), *The manipulation of human behavior.* New York: Wiley, 1961. Pp. 51–95.

Lindsley, D. B. Common factors in sensory deprivation, sensory distortion, and sensory overload. In P. Solomon et al. (Eds.), *Sensory deprivation.* Cambridge: Harvard Univer. Press, 1961. Pp. 174–194.

Malitz, S., Wilkens, B., & Esecover, H. A comparison of drug-induced hallucinations

with those seen in spontaneously occurring psychoses. In L. J. West (Ed.), *Hallucinations.* New York: Grune & Stratton, 1962. Pp. 50–63.

MENDELSON, J., SOLOMON, P., & LINDEMANN, E. Hallucinations of poliomyelitis patients during treatment in a respirator. *J. nerv. ment. Dis.,* 1958, **126,** 421–428.

MURPHY, D. B., MYERS, T. I., & SMITH, S. Reported visual sensations as a function of sustained sensory deprivation and social isolation. *USA Leadership HRU Draft res. Rep., Presidio of Monterey,* 1962. (Pioneer VI)

MYERS, T. I., & MURPHY, D. B. Reported visual sensation during brief exposure to reduced sensory input. In L. J. West (Ed.), *Hallucinations.* New York: Grune & Stratton, 1962. Pp. 118–124.

POLLARD, J. C., UHR, L., & JACKSON, C. W., JR. Some unexpected findings in experimental sensory deprivation: The psycho-pharmacologic interaction of a placebo-potentiated suggestion. Paper read at American Psychiatric Association, St. Louis, 1963. (a)

POLLARD, J. C., UHR, L., & JACKSON, C. W., JR. Studies in sensory deprivation. *Arch. gen. Psychiat.,* 1963, 8, 435–454. (b)

ROSENZWEIG, N. Sensory deprivation and schizophrenia: Some clinical and theoretical similarities. *Amer. J. Psychiat.,* 1959, **116,** 326–329.

ROSSI, A. M., STURROCK, J. B., & SOLOMON, P. Suggestion effects on reported imagery in sensory deprivation. *Percept. mot. Skills,* 1963, **16,** 39–45.

RUFF, G. E., LEVY, E. Z., & THALER, V. H. Factors influencing the reaction to reduced sensory input. In P. Solomon et al. (Eds.), *Sensory deprivation.* Cambridge: Harvard Univer. Press, 1961. Pp. 72–90.

SCHEIBEL, MADGE E., & SCHEIBEL, A. B. Hallucinations and brain stem reticular core. In L. J. West (Ed.), *Hallucinations.* New York: Grune & Stratton, 1962. Pp. 15–35.

SHURLEY, J. T. Mental imagery in profound experimental sensory isolation. In L. J. West (Ed.), *Hallucinations.* New York: Grune & Stratton, 1962. Pp. 153–157.

SILVERMAN, A. J., COHEN, S. I., SHMAVONIAN, B. M., & GREENBERG, G. Psychophysiological investigations in sensory deprivation. *Psychosom. Med.,* 1961, **23,** 48–61.

SOLOMON, P., KUBZANSKY, P. E., LEIDERMAN, P. H., MENDELSON, J. H., TRUMBULL, R., & WEXLER, D. (Eds.) *Sensory deprivation.* Cambridge: Harvard Univer. Press, 1961.

SOLOMON, P., & MENDELSON, J. Hallucinations in sensory deprivation. In L. J. West (Ed.), *Hallucinations.* New York: Grune & Stratton, 1962. Pp. 135–145.

VERNON, J., MCGILL, T. E., & SCHIFFMAN, H. Visual hallucinations during perceptual isolation. *Canad. J. Psychol.,* 1958, **12,** 31–34.

VERNON, J., MARTON, T., & PETERSON, E. Sensory deprivation and hallucinations. *Science,* 1961, **133,** 1808–1812.

VOSBERG, R., FRASER, N., & GUEHL, J. Imagery sequence in sensory deprivation. *AMA Arch. gen. Psychiat.,* 1960, **2,** 356–357.

WALTERS, R. H. & HENNING, G. B. Social isolation, effect of instructions and verbal behavior. *Canad. J. Psychol.,* 1962, **16,** 202–210.

WALTERS, R. H., & QUINN, M. J. The effects of social and sensory deprivation on autokinetic judgments. *J. Pers.,* 1960, **28,** 210–219.

WEST, L. J. (Ed.) *Hallucinations.* New York: Grune & Stratton, 1962.

WEXLER, D., MENDELSON, J., LEIDERMAN, P. H., & SOLOMON, P. Sensory deprivation: A technique for studying psychiatric aspects of stress. *Arch. neurol. Psychiat.,* 1958, **79,** 225–233.

ZUBEK, J. P., AFTANAS, M., HASEK, J., SANSOM, W., SCHLUDERMANN, E., WILGOSH, L., & WINOCUR, G. Intellectual and perceptual changes during prolonged perceptual deprivation: Low illumination and noise level. *Percept. mot. Skills,* 1962, **15,** 171–198.

ZUBEK, J. P., AFTANAS, M., KOVACH, K., WILGOSH, L., & WINOCUR, G. Effect of severe immobilization of the body on intellectual and perceptual processes. *Canad. J. Psychol.,* 1963, **17,** 118–133.

ZUBEK, J. P., PUSHKAR, D., SANSOM, W., & GOWING, J. Perceptual changes after prolonged sensory isolation (darkness and silence). *Canad. J. Psychol.,* 1961, **15,** 83–101.

ZUBEK, J. P., WELCH, G., & SAUNDERS, M. G. Electroencephalographic changes during and after 14 days of perceptual deprivation. *Science,* 1963, **139,** 490–492.

ZUCKERMAN, M., ALBRIGHT, R. J., MARKS, C. S., & MILLER, G. L. Stress and hallucinatory effects of perceptual isolation and confinement. *Psychol. Monogr.,* 1962, **76**(30, Whole No. 549).

ZUCKERMAN, M., & COHEN, N. Is suggestion the source of reported visual sensations in perceptual isolation? *J. abnorm. soc. Psychol.,* 1964, in press.

ZUCKERMAN, M., LEVINE, S., & BIASE, D. V. Stress responses in total and partial perceptual isolation. *Psychosom. Med.,* 1964, in press.

(Received July 26, 1963)

Psychological Bulletin
1964, Vol. 62, No. 1, 21–35

PERCEPTION IN SOVIET PSYCHOLOGY

HERBERT L. PICK, JR.[1]
Institute of Child Development, University of Minnesota

The relation between experimental work in perception in the Soviet Union and the orthodox philosophy of the Soviet state is discussed. It is pointed out that both the philosophy derived from Lenin and Marx and the needs of the state exert an influence on the nature of the problems studied and the type of theory developed. Studies in the areas of tactual, auditory, and visual perception are reviewed. These are considered against the background of Leont'ev's motor-copy theory of perception, currently the most systematic Soviet approach to perception. Certain other studies in visual perception conducted in the context of Uznadze's theory of set, as well as some mathematical approaches to perception not fitting into any general theoretical frame, are also briefly reviewed.

Three prominent issues are apparent in Soviet work in the field of perception: an epistemological concern, an emphasis on response processes, and a pragmatic interest. It is the purpose of the present review to describe several areas of research pursued by Soviet psychologists interested in perception. The discussion will attempt to point out the manifestations of these trends and at the same time to be detailed enough to present some essence and perhaps more flavor of the research.

The epistemological current in Soviet psychology derives from the accepted state philosophy, Marxist materialism—in this particular case Lenin's elaboration of Marx as it is applied to sensation and perception. In brief the Soviet philosophical position might be characterized as follows: The materialist position argues for the existence of the real world and the primacy of matter against any kind of idealist philosophy. Further, Lenin in an emotional polemic claims that we know the real world as it actually exists. This knowledge is obtained through our sense organs and our brain mirrors the world (Lenin, 1927). Lenin's theory of the reflection of the real world has been termed by some a copy theory and is discussed in detail by Wetter (1958).

The Soviet psychologists' assumption that perception mirrors the real world leads them to be concerned with images and their formation in consciousness. When one examines the procedure by which the existence and qualities of an image are determined, it is sometimes easily translated into operations. For example, in a study of the role of eye movements in young children's perception (Zinchenko, Chzhi-tsin, & Tarakanov, 1962), the authors mention the "adequacy" of the image of an object but they use a multiple-choice recognition procedure and error score to measure this.

One implication of the copy theory is that we do not have to add anything to the information obtained via sense organs. Central addition to sensory information in order to obtain true knowledge smacks of idealism and is rejected. Idealism (as opposed to materialism)

[1] The author is indebted to A. N. Leont'ev, A. V. Zaporozhets, the members of the Department of Psychology of Moscow State University, and to the many Soviet psychologists who helped him become familiar with their work during the academic year 1959–60. He is further indebted to William Charlesworth and Anne D. Pick for their critical and helpful reading of the draft of this paper.

here refers to any possible sense in which ideas define or determine reality. The concept that association or integration is imposed on stimulation by the central nervous system might lead to the possibility that ideas define reality. On the other hand, kinesthetic feedback and other combined sensory inputs from several modalities are emphasized by Soviet psychologists (Leont'ev, 1959b). This fact suggests that basically the Soviet psychologists may be rejecting successive association and Gestalt organization for simultaneous association.

The importance of kinesthetic feedback or response-produced stimulation in Soviet theory seems to derive from a happy coincidence of certain philosophical points in Marx and Engels and some theoretical ideas of Sechinov and Pavlov. Marx in his *Theses on Feuerbach* implied that truth is determined in activity, in particular, that sensory activity and practice, not passive contemplation, determine reality (Aikin, 1956, p. 193).

The theoretical ideas of Sechinov come from his observations of behavior and subsequent speculations (cf. Leont'ev, 1959a) to the effect that sense organs, exemplified by the hand, obtain information by actively exploring the objects in the environment. These suggestions are further reinforced by Pavlov's interest in the orienting response and in the kinesthetic analyzer.

The very real requirements of a rapidly industrializing country and the emphasis on action as exemplified by Marx's *Theses on Feuerbach* (Aikin, 1956, p. 195) give to Soviet psychology quite a pragmatic cast.

The considerations outlined above, the materialist assumptions and the response emphasis, are embodied in the most important current Soviet theory of perception—one developed by Leont'ev (1959a). In this theory the organism is considered to make a reflex response to a stimulus. The combination of receptor stimulation and the reflex feedback form the basis of the perception. In one case (Leont'ev, 1959b), the stimulus is conceived as evoking an orienting response to which the organism is initially insensitive. Then by a classical conditioning procedure the attention of the subject is attracted to the reflex state itself and, in this mediated way, the subject is able to respond to the original stimulus. An example of this is the stimulation by a light of the skin of a blindfolded subject. Light with appropriate heat filters had been shown to elicit a physiological skin reaction to which subjects ordinarily do not respond. When the light is consistently followed by an electric shock, subjects slowly come to be able to anticipate the shock although they do not know the basis of the anticipation. In another case, tactual stimulation and kinesthetic feedback combine to yield the perception of specific objects in active touch (Anan'ev, Vekker, Lomov, & Yarmolenko, 1959). Here, it seems that the physical object constrains the hand to reproduce the contour of the object, and feedback from this motion forms the basis of perception of the object. The initial interaction of the hand and object with its correlative tactual stimulation forms the basis of the perception of such qualities as temperature, hardness, elasticity, etc. But even these depend to a certain extent on feedback from the hand movement.

In the first of the above cases (the stimulation of the skin by a light), Leont'ev's idea was that a reflex response is made to a stimulus and the subject becomes sensitive to this response and not to the conditioned stimulus. Leont'ev employs this type of process to explain "sensation." In the second case, the response that we are perceiving is a reproduction (by our

own movements) of the physical stimulus. This general process forms the basis of Leont'ev's theory of perception. We in some sense reproduce the physical object by our movements. The combination of kinesthetic feedback and initial stimulation serve as the basis of our perception. This is a motor theory of perception, but a motor theory which involves copying the physical stimulus.

It is easy to point out difficulties for the theory, the same difficulties with which our motor theorists of cognition and perception have been faced. Instead, let us consider three sets of investigations which have been instigated by this view of perception. One set is on tactual perception, the second on the development of visual and tactual perception, and the third on auditory-pitch perception.

The studies of tactual perception (Anan'ev et al., 1959) generally involved asking a blindfolded subject to explore an object by hand and then either to make a drawing of it or to recognize it among several objects. Particular attention was paid in many of their investigations to the manner in which the subject explored. This, of course, is a manifestation of the interest in response processes described above. The question of how the movements of the hand provide the basic information for perception of the object is important here.

Very often time-motion analyses of the hand movements of the subjects exploring the objects were made. From motion picture films a "tsiklogram" was constructed consisting of a plot over time of the positions of the hands and fingers. From this the speed of movements along the contour could easily be analyzed. Such analyses carried out for the separate fingers, exploring an object bimanually, indicated that the different fingers explored at different rates and stopped at different times. Pauses typically occurred at corners. A sample (Anan'ev et al., 1959) of the results of such an analysis of a subject, who has been asked to explore a shoelike cutout form, indicated that all the fingers of the right hand were in motion more of the time than they were at rest during the exploration, while for the left hand the index finger and fourth finger were at rest more than they were moving. These relations changed somewhat upon repeated exploration of the same object. In spite of such methods which easily lend themselves to quantification, the investigators presented very little quantitative data and appeared to be more interested in a qualitative analysis of the process of tactual exploration.

The picture presented (Anan'ev et al., 1959) of the bimanual exploration of an object (under directions to the subject to reach behind a screen and explore an object so that he will be able to make a drawing of it) starts with movements of both hands in the air or on a table along the sagittal axis of the body until contact with the object is made. The hands then slide lightly over the surface of the object to the upper edge. These movements are said to be regulated mainly by stimulation from the middle finger. When the hands get to the far point of the object, they stop for a moment. All of these movements are said to be orienting in nature. Reproductions made after this initial exploration are poor in detail but good in general form although often elongated along the vertical axis. It is noted that such reproductions are made in the order that the exploration occurred and subjects have great difficulty in changing the order of reproduction.

The subsequent exploring motions consist of both gross hand and finger movements and micro finger movements of a millimeter or two, the purpose of the latter being interpreted as mainte-

nance of sensitivity. The function of the gross movements is said to be synthesis of the multitude of cues being perceived. The basis of this interpretation is that the subject having made these movements is then able to draw a picture of the object starting from any point of the contour. The authors suggest that the spatial-temporal image has developed into a spatial image.

The form of the gross exploring movements is more or less determined by the shape of the object. The trajectory of the movements corresponds in general to the shape. The optimal speed is between 5 and 10 centimeters per second. The movements are not even, but broken into a series of small sections. The pauses occur at apex angles and changes of direction of movement. The two hands move and pause at the same time when exploring symmetric objects. But if one hand reaches a stopping point it pauses until the other catches up. If the object is not bilaterally symmetric the movements of the hands become asynchronous and alternating—one hand moving and the other acting as a reference point for the exploration.

In a fashion as descriptive as the above, many aspects of touch were analyzed: monomanual touch, touch with various fingers immobilized, passive touch, touch with instruments including artificial limbs, touch in the physically handicapped, touch in the process of manual labor, etc.

One hypothesis presented in this monograph, namely, that the nature of the interaction between receptor and object determines the properties of the perception, has been further investigated by Vekker and Lope (1961). These experimenters moved geometric forms with sides of various compositions across a subject's passive finger. They report a curvilinear relation between the roughness and estimated length. The smoothest and very roughest lengths were estimated as longer than similar lengths of intermediate roughness. The same results were obtained with active touch using a single finger. Duration of movement is positively correlated with estimated length of object; if the duration is doubled, estimated length is increased by 50%.

Another aspect of the use of information from moving stimuli reported in the same paper is that adequacy of perception of stimuli moving on a particular part of the skin is positively correlated with the absolute threshold of that part of the skin. In this connection it should also be mentioned that Mileryan and Tkachenko (1961) reported some work in the tradition of classical threshold psychophysics. They gave subjects 12 days of practice in spatial discrimination of tactual stimuli. Their report indicates that thresholds decreased markedly for the practiced section of skin and the increased sensitivity generalized to neighboring areas decreasing as the areas became more remote. There was, however, complete bilateral transfer of improved discrimination.

Anan'ev and his colleagues suggested that tactual perception is more primitive ontogenetically than visual perception. They repeat rather often the assertion that touch teaches vision. No evidence is presented for this assertion, but the following investigations of Zinchenko are concerned with the development of tactual and visual perception and with cross-modal perception.

Zinchenko (1957; Zinchenko, Lomov, & Ruzskaya, 1959) was struck by the obvious fact that gross eye movements are not necessary for perception of objects by adults. This posed an immediate problem for a motor theory of perception. Moreover he observed that although the hand was the organ par excellence for the motor-copy theory, it

could also, under unusual conditions, perceive while remaining quite passive.

Reviewing results of other investigators (both Soviet and Western) Zinchenko (1957) concluded that: (*a*) Young children (3 years old) do not initially visually differentiate geometric forms presented to them. They orient towards the forms tactually and only when the hand movements around the contour of the objects become orderly do the eyes begin to track the hands. Tactual differentiation precedes visual differentiation. (*b*) Tactual identification of unfamiliar objects by adults is quite slow, requiring much preliminary exploration. It becomes almost instantaneous as the object becomes very familiar. (*c*) Visual inspection of new objects also involves many large movements initially. With repeated inspection these movements become more abbreviated and conform more closely to the contour of the object, finally dropping out completely. On the basis of such analyses of previous work, Zinchenko felt that visual and tactual perception were not as different as had been previously considered.

These considerations inspired an additional series of studies on the comparative development of touch and vision (Chzhi-tsin, Zinchenko, & Ruzskaya, 1961; Lavrent'eva & Ruzskaya, 1960; Tarakanov & Zinchenko, 1960; Zinchenko, 1960; Zinchenko, Chzhi-tsin, & Tarakanov, 1962; Zinchenko & Ruzskaya, 1960a, 1960b).

The majority of these experiments were concerned with the perception of forms both within one sense modality and cross modally. The forms used were rather amorphous two-dimensional cutouts adapted from those used by Gaydos (1956). The stimulus was presented to the subject visually or tactually and required the subject to recognize it when presented either visually or tactually along with two other forms. The subjects were preschool children aged 3 to 7. Quantitative observations included percentage of errors in recognition and amount of time spent in the recognition part of the task. Qualitative observations were made of the type of exploratory movements, both hand and eye, which occurred as the children performed the task.

In general, errors decreased with age. Wherever big improvement in error score occurred it was interpreted in light of the changes in hand and/or eye movements which occurred along with it. Thus eye movements of 3–4 year olds in the visual-visual combination did not follow the contour of the form, but consisted of saccadic movements almost entirely inside the contour. These children made approximately 50% errors. The 4–5 year olds made 28.5% errors and their eye movements conformed much more closely to the contour.

These experiments while quite interesting in principle are somewhat disappointing in execution. Control for order and practice effects appears to be lacking. Sometimes much is made of small differences in percentage of errors which probably would not be significant if statistical analyses were made. The results are suggestive nevertheless and the percentage of errors for the different tasks is shown in Table 1.

The next experiments in the series were designed to study procedures for improving performance on this type of task. Lavrent'eva and Ruzskaya (1960) tried to improve cross-modal matching in children by making the task one of simultaneous matching rather than successive, on the supposition that the poor performance of the younger children might be due to the memory requirements of the successive situation. Children were shown one stimulus in one sense modality and asked to recognize it simultaneously from a group of

TABLE 1

Percentage of Errors in Recognition of Perceived Figures

Condition of initial perception	Condition of recognition	Percentage of errors at age: 3–4	4–5	5–6	6–7
Visual	Visual	50.0	28.5	0	2.5
Visual	Tactual	—[a]	73.0	34.0	23.2
Tactual	Tactual	47.7	42.3	25.0	23.1
Tactual	Visual	70.5	42.3	38.5	40.4

[a] Subjects unable to respond meaningfully to this task.

three through the other modality. To the surprise of the investigators, performance was worse on this simultaneous task than on the previous successive tasks, particularly for the younger children. The relative failure of the younger children is attributed to their tendency to orient towards the recognition stimuli and not towards the single stimulus to be matched. This conclusion is supported by data of exploratory time spent with the different stimuli. The children who spend practically no time examining the stimulus to be matched can hardly develop an image of it. The older children do spend as much as one third of their total time exploring the single stimulus and consequently do relatively better.

The second procedure used to improve performance was reported by Tarakanov and Zinchenko (1960). They had one group of children examine a stimulus tactually and visually for 10 seconds, then recognize it from among three, tactually by one subgroup and visually by another. A second group was required to fit the initial stimulus form into one of three holes which matched it in size and shape. In this case, the second group was markedly superior to the first group on both visual and tactual recognition tasks. The superiority of the second group is attributed to the fact that the practical task required of the subjects insured that the proper aspects of the stimuli were attended to. The authors suggest that generally children of this age learn more when performing a practical activity than when instructed to remember something for subsequent use.

Finally, a straightforward attempt to train children to perform such a matching task was made by Chzhi-Tsin, Zinchenko, and Ruzskaya (1961). Children were given three successive stages of training. In the first stage the subject was asked to follow, by eye, the movement of a pointer with which the experimenter traced around the contour of the stimulus form. In the second stage the subject was shown how to grope (with his fingers) around the contours of the forms presented tactually. In the third stage the subject was asked to move his own finger around the contour of the original single stimulus and to follow his own movements by eye. Error scores indicated that performance improved at each training stage but the biggest improvement occurred at the first stage. The authors concluded that requiring the subject to visually follow the experimenter's pointer moving around the contour, is the most effective procedure and it remains effective longest. However, no controls were included for order and practice effects of the successive stages of training so the conclusions must be accepted with extreme caution.

In reviewing the results of several of these experiments, Zinchenko and Ruzskaya (1960b) noted that, under

various conditions, visual perception proceeds sooner and more accurately than tactual, and that tactual perception interferes in some cases with visual. These results forced them to reject their initial formula that the hand teaches the eye.

The experiments considered so far have all involved receptor systems in which the receptor organs and the organ "copying" the stimulus were one and the same, the eye or the hand. A more critical demonstration of the generality of their motor-copy theory would be in a modality of perception for which such correspondence was not the case. Leont'ev and his colleagues have conducted a series of experiments on pitch perception with just such an aim in view. The ear, of course, is the receptor organ for auditory stimuli, but if any motor system is involved in copying the physical stimulus it would be the voice musculature. Leont'ev approached the study of pitch perception by trying to isolate it from its dependence on timbre.

The initial experiment was the determination of the difference threshold for pitch, first in the classical manner, that is, with pure tones, and again with the pitch-timbre correspondence distorted by adding specific harmonics to a base tone (Gippenreiter, 1957, 1960). Both threshold determinations involved presentation to the subject of a series of pairs of tones. Subjects were asked to judge which of the two tones was higher in pitch.

The thresholds determined in the classical manner ranged from 5 to 135 musical cents. The timbre-distorted thresholds (tdt) showed a sharp increase in most cases over the classical thresholds. Three groups of subjects were distinguished: a small minority (13%) whose thresholds did not change; a second group (57%) whose thresholds deteriorated moderately, for example, by a factor of three or four; and a third group (30%) whose thresholds for all intents and purposes deteriorated so as to be unmeasurable (specifically being greater than 1,200 musical cents). There also appeared to be little correlation between threshold for pure pitch and tdt's. This fact is interpreted as supporting a hypothesis that there are two separate systems involving the perception of pitch.

The next step, in investigating this problem, involved application of the motor-copy theory by studying the relation between ability to vocalize tones and ability to discriminate pitch. Three small studies were conducted (Gippenreiter, 1958, 1960). The first began with a replication of the tdt determinations with a new group of subjects, then tdt values were obtained requiring the subject to vocalize the tones before making the judgments as to which tone was of higher pitch. This vocalizing procedure resulted in a sharp decrease in threshold over the values obtained without vocalization. Some subjects were retested under instructions to remain silent. Their thresholds reverted to the initial high values. The second study was a determination of tdt's for people who had narrow ranges of accurate vocalization. For tones within this range these subjects had very small difference thresholds. Finally in the third small study it was determined that the overall correlation between accuracy of vocalization and tdt was .83. Such data provide correlational evidence of a close association between vocalization and discrimination but they do not demonstrate the nature of this association.

Further support for this association was adduced on the basis of experiments in which thresholds were determined while the subject was asked to intonate a different note from the one presented (Ovchinnikova, 1958, 1960b).

Subjects with both small and large tdt's were tested in this manner. They were required to sing a given note while tdt determinations were made at a different range on the scale. Subjects with large tdt's were not much affected by singing. Their pitch thresholds were so bad to start with under these timbre-distorted conditions that they could not get worse. Of the 18 subjects in the groups with small tdt's, 12 showed a marked increase in threshold when the determination was made while intonating a note. The other 6 showed little change in threshold. Upon questioning, 4 of these latter reported having had a good deal of experience singing in choruses. The inference from these results, that motor involvement causes the deterioration of threshold, is weakened by the possibility that the sound generated by the subject himself interfered with the task of discrimination. In addition it could be argued that the subject has to attend to two tasks and becomes less efficient. The interference criticism is eliminated in a purer experiment described below. The present experiment along with the other evidence accumulated is certainly supportive of the hypothesis of vocal-motor involvement in pitch perception.

On the basis of the data, Leont'ev and his colleagues hypothesized that there are indeed two functional systems for auditory perception (Gippenreiter, 1960; Leont'ev & Gippenreiter, 1959). They suggested that these perceptual systems develop early ontogenetically and independently—the one depending on speech training, the other on musical training. In some cases the system depending on speech training develops normally and replaces or suppresses the other system in which pure pitch discriminations are important. In languages such as Vietnamese where the pitch itself carries meaning, there should be no chance for such replacement or suppression. With these considerations in mind the original pure tone and tdt experiments were carried out on a group of 20 Vietnamese college students. Table 2 abstracted from Leont'ev and Gippenreiter (1959) presents a comparison of the results obtained from Russians and Vietnamese.

Having gone about as far as possible to support the motor-copy theory on the basis of normative data, the next step was to try to experimentally manipulate the precision of perception of this timbre-distorted pitch—in short, training experiments. Two methods of training were tried (Ovchinnikova, 1959a, 1959b, 1960b). The first was a sensory training method consisting essentially of correcting the subject during the same psychophysical procedure that was used to test thresholds. Before and after this training, tdt's were determined as were thresholds for pure tones. This training, carried out on nine subjects, resulted in a decrease in threshold for the pure tones but no significant generalization to the timbre-distorted tones.

The second method involved motor training. Here subjects were presented with single pure tones and were asked to sing them. The experimenter verbally corrected the subject's performance,

TABLE 2

PERCENTAGE OF SUBJECTS WITH INDICATED CHANGE OF TDT

Group	Unchanged	Increased by a factor of 2	Increased by a factor 2–4	Increased by a factor of 5 or more
Vietnamese	50	25	25	0
Russian	26	20	13.5	40.5

coaching him on to the correct tone. All nine subjects given this motor training improved in their ability to intonate these tones accurately and all improved considerably in their tdt's. Those who improved most reported singing the notes to themselves. Those who improved least did not sing the notes. When the latter subjects were instructed to sing the notes aloud before making discriminative judgments, their tdt's dropped to a much lower level.

The authors interpreted this as suggesting that, on the basis of reinforcement, an association is established between the pitch of the sound and a particular activity of the vocal chords. This activity then serves as an indicator of the pitch. Such an interpretation implies a two-level theory of perception (which is, of course, implicit in the whole motor-copy hypothesis) in which an apparently lower-level reflex response occurs which forms the basis of the conscious discrimination of pitch.

Independently of these investigations, Chistovich and her colleagues working in the Pavlovian Institute in Leningrad have also studied the role of vocalization in auditory perception (Chistovich, Alyakrinskii, & Albul'yan, 1960; Chistovich, Klass, & Alekin, 1961). In general her approach has been to have subjects repeat (identify) or discriminate (same-different judgments) sounds presented to them. Accuracy of both repetition and discrimination are better for languagelike sounds than for random tones. Quantitative analysis of the results was carried out in terms of information transmitted.

To return to Leont'ev, his model, as it so far has been presented, can be considered to involve a stimulus exciting a receptor organ (sensory link) and a reflex response (motor link), the latter serving as a mediator. The final two experiments in the series replace the sensory link on the one hand (Leont'ev, 1960) and the motor link on the other hand (Ovchinnikova, 1960a).

The auditory sensory link was replaced as follows: Instead of sounding a note, a vibratory stimulus was applied to the skin of the subject. Vibratory stimulation, according to Leont'ev, has, like sound, two often confused parameters, specifically frequency and intensity. That is, increasing the intensity of the vibration typically results in a decrease of perceived frequency and conversely. A noiseless vibrator was applied to the index finger in this study. At first, frequency difference thresholds were obtained for vibrations of the same intensity, then with the intensity of one stimulus half that of the other. In the latter case the difference threshold was increased by a factor of three to five. The subjects were then trained to intonate notes of the same frequency as the vibratory stimuli. Slowly, on the basis of verbal reinforcement from the experimenter they became able to match the frequency of the stimuli with their voices. Difference thresholds were redetermined and differential sensitivity to these amplitude-distorted vibrations had increased.

In the experiment in which the motor link was replaced, the auditory system was utilized as in the earlier experiments on pitch. But here when a note was sounded instead of singing it as in the previous auditory motor training procedure, the subjects were asked to press a key with a pressure defined by an arbitrary scale which linearly related pressure to frequency. The experimenter verbally corrected the subject's pressure. The experiment was carried out with three subjects who initially had large tdt's. After numerous (25–33) training sessions the subjects learned this arbitrary scale. Their tdt's

were then redetermined and showed large increases in sensitivity. Furthermore if the subject's hand was involved in some extraneous task his tdt's reverted back to the initial high levels.

These are rather clever experiments with which the series is ended, but the last experiment contradicts the motor-copy hypothesis. In particular, the subject does not have to produce a tone with his muscle pressure in order to differentiate pitch. Interestingly the apparatus was wired so that the subject's pressure produced a tone via an audio oscillator. However, the subject did not hear the tone. The experimenter monitored the pressure by watching the tone on an oscilloscope. It would seem as if the copy part of the motor-copy hypothesis has been shown to be unnecessary in this experiment.

There have been no other such systematic, theoretically integrated research programs carried out in auditory perception by Soviet psychologists although the work of Chistovich referred to briefly above approaches this in some respects and is of very high quality. Chistovich's colleague, Maruseva (1959a, 1959b), has extensively studied the functioning of the auditory system both in children and adults. This work has also been concerned with determination of sensitivity and with possible ways of increasing sensitivity but has been done from a more strictly classical psychophysiological point of view. There has been a recent attempt to relate individual differences in auditory perception to occupation (Kalikinskiĭ, 1961a, 1961b) and a study of the development of speech comprehension (Liamina, 1960), but such work seems to be sporadic and of questionable quality.

Some work on visual perception has been described in various connections but Soviet contributions in this area have generally not been systematic or profound. General discussions of perception such as Anan'ev's (1960) adopt a generally empirical approach to visual perception. Within this point of view some of the classical problems of space perception are interpreted from a Pavlovian model of learning. Thus some of the classical cues like convergence, accommodation, etc., become conditioned stimuli for the perception of depth. The Soviet interpretation would like, as usual, a reflex response to form the initial basis of depth discrimination. Such an argument can be made for convergence and accommodation. Anan'ev (1960) finds it difficult in the case of retinal disparity. But by a rather abstruse route he also attempts to put this on a reflex basis (pp. 154–160). He cites experiments conducted in his laboratory which indicate that eye dominance shifts with change of viewing distance. He then suggests that this shift of eye dominance is the conditioned reflex to a change of distance. Even granting this claim, it is not clear to this reviewer how the function of retinal disparity in depth perception is thereby explained.

The problem of constancy has been approached more or less systematically by Georgian psychologists. They have been working in the context of Uznadze's theory of set. Many psychological phenomona are interpreted as manifestations of established sets. A thorough exposition of this point of view is presented by Prangishvili and Khodzhava (1958). In fairness it should be stated that the proponents of this school are concerned both with how to establish a set and its subsequent effects. However, it appears to be a rather naive approach and does not offer very satisfying explanations. When the concept of set is applied to the problem of constancy, we find, for example, Natadze (1961) simply suggesting that all theories of constancy

demand some sort of evaluation of the situation by the subject and since this goes on unconsciously it may be simply that the subject is set for a particular situation. The same author (Natadze, 1960) describes an apparatus similar to that of Holway and Boring (1941) for varying the amount of information (secondary spatial cues) available to the subject as to the distance at which he is making size judgments. When he is thus set for the actual distance, constancy is good. Working in the opposite direction, Adamashvili (1960) creates a false set and shows how this distorts size judgments. Part of his report concerns the way in which independent variation of size and distance of stimuli can effect the set established in the subject.

Bzhalava (1962) has applied the concept of set to figural aftereffects of the Köhler-Wallach and Gibson types. The line of reasoning as far as the Köhler-Wallach aftereffects are concerned is that fixation of the inspection figure induces a set to see a particular size figure in a particular part of the visual field. When a smaller or larger test figure is subsequently introduced, a contrast illusion results in the test figure looking even smaller or larger than it actually is. Experiments are reported in which such contrast effects are obtained when the inspection and test figures differ considerably in shape. The analysis is provocative, but all the variations of Köhler-Wallach aftereffects cannot be so simply handled, and a critical demonstration that a general set can handle these phenomena depends on a careful analysis of the precise shapes perceived upon presentation of the test figures. Such an analysis is not presented. Finally "set" so far has been too vaguely defined to constitute an explanatory concept.

There are isolated studies on other aspects of visual perception to be found in Soviet psychology. For example, Dymerskiĭ (1960) completed a dissertation on the perception of spatial-temporal relations. He was primarily interested in the cues used by pilots in landing airplanes. Using experimental studies of the perception of distance by moving observers, the practical experience of pilots, and mathematical analysis, he established a general relation between perceived distance and various components of angular acceleration. In addition, it was suggested that the optimum place a pilot should look in landing an airplane depends on the difference between the change of angular velocity due to changing altitude and the change in angular velocity due to change in direction of movement with respect to the ground surface.

Another example is the work of Fonarev (1959) who investigated the presence of conjugate eye movements in neonates. He found that in the absence of visual stimulation (very low uniform illumination) the majority of neonates' eye movements were conjugate. However, introduction of a stimulus inhibits this conjunction and the majority of eye movements become disjunctive.

An interest in eye movements is quite congruent with the motor-copy theory as suggested previously and there is a large amount of interest in eye movements but no systematic research programs. Luriya, Providina-vinarskaya, and Yarbus (1961) present an interesting study of a case of optic ataxia in which the patient could fixate and track a moving stimulus but was unable to shift point of fixation at will. Zinchenko (1957) was interested in the possibility of conditioning exploratory eye movements. He found that, under certain conditions of practice, eye movements anticipated the onset of a light stimulus if an appropriate signal preceded it. Khomskaya (1962) has

studied eye movements from the point of view of determination of the necessary stimulus conditions for movement. She recorded eye movements by means of photoelectric cells activated by light reflected from the eye. Subjects were asked to perform four tasks: (*a*) to look back and forth between two points separated by a visual angle of 30 degrees, (*b*) to follow a moving light which moved back and forth at rates of .32–2.08 cycles per second, (*c*) to repeat Task *a* to see if fatigue effects occurred, and (*d*) to track a moving light by memory, that is, to continue tracking movements after the moving light was extinguished. In Task *a* saccadic movements occurred between the two points with short fixations at the two points. The frequency of the movements was very constant for each subject but large individual differences were noted. In Task *b* eye movements tracked the light quite closely at low frequencies but at high frequency movements became saccadic again. In Task *c* there were no differences from Task *a*. And in Task *d* as soon as the actual moving light was extinguished, the subject's eye movements became saccadic.

One final significant trend in the Soviet study of perception is the application of mathematical methods. The dominant figure in this work is E. N. Sokolov, better known for his work on the orienting reflex. Sokolov and Mikhalevskaya (1961a, 1962) have applied mathematical techniques to evaluation of the effectiveness of near-threshold stimuli. In one case they used Bayes theorem to estimate the conditional probability that the stimulus had occurred given that a particular EEG response was present. In another case, an investigation of sensitivity to light, they empirically determined the general relation between reaction time and intensity of light stimulus. After fixing confidence bounds for the function, they were able to estimate the probability that specific stimuli were above, at, or below threshold, given a particular reaction time, threshold being defined as that value of stimulus perceived a certain percentage of the time.

A few years ago Sokolov became concerned with scanning mechanisms which might be used for letter recognition. His approach (Arana, 1961; Sokolov, 1960) was to look at an ideally efficient form of scanning behavior (one which minimizes area scanned) and to compare this with actual behavior of subjects. Letters of the alphabet were to be individually scanned until recognition occurred. To make the experimental situation one of sequential scanning a tactual task was used with vision precluded.

A probability model for ideal scanning was constructed using successive application of Bayes theorem. That is, given that the *n*th square is filled, what is the probability that the letter is A, B, C . . . ? And given the new probabilities, what is ideally the next square to scan? Analyses of situations employing various numbers of letters and fuzziness of letters have been made. A subject's behavior in such situations show a gradual approach to the most efficient scanning behavior for the set of letters used. Using a large number of letters has the effect of increasing the number of superfluous scanning movements. But gradually increasing the number of letters after efficient scanning behavior has developed results in preservation of the scanning strategies developed by the subject.

In another approach, Sokolov and Mikhalevskaya (1961b) elaborated a probability and information analysis for discrimination situations, much like the analysis of Garner and Hake (1951). There is a lively interest in the application of information theory to

psychology in the Soviet Union. There conceivably are difficulties in integrating an information-theory approach with Lenin's copy theory. An information-theory model for perception would not seem to require that a "copy" of the stimulus arrive in the head since obtaining information strictly speaking only involves reducing the uncertainty in designating which stimulus has occurred. The reviewer, however, has seen no Soviet commentaries on any such difficulty so perhaps it is not the problem he imagines.

This final area of Soviet research is apparently one in which Soviet philosophy and the needs of the state have not imposed their influence. But this is not usual and, in general, it has been suggested that the motor-copy theory and the resarch inspired by it have sanction, if not the enouragement of orthodox Soviet philosophy. The theory and research reflect as well the emphasis on response processes derived from the Russian physiological tradition. Several aspects of the research, in visual perception (aircraft landings), in auditory perception (occupational differences), and tactual perception (touch in manual labor) have a definite practical orientation. Soviet psychology is a relatively small enterprise in comparison with American psychology. It is to be hoped that as it increases in size and importance, and with a general liberalization in the Soviet Union, some of the constraints will evaporate and Soviet psychologists will apply their creativity and industriousness in many more directions.

REFERENCES

ADAMASHVILI, N. G. Otrazhenie v fiksirovanoĭ ustanovke sootnosheniya velichinў i otdalennosti predmetov. [The reflection of the interrelation of size and distance under conditions of set.] *Vop. Psikhol.*, 1960, 6(3), 33–45.

AIKIN, H. D. (Ed.) *The age of ideology.* New York: New American Library of World Literature, 1956.

ANAN'EV, B. G. *Psikhologiya chuvstvennogo poznaniya.* [Psychology of sensory cognition.] Moscow: Publishing House of the Academy of Pedagogical Sciences RSFSR, 1960.

ANAN'EV, B. G., VEKKER, L. M., LOMOV, B. F., & YARMOLENKO, A. V. *Osyazanie v protsessakh poznaniya i truda.* [Touch in the processes of cognition and work.] Moscow: Publishing House of the Academy of Pedagogical Sciences RSFSR, 1959.

ARANA, L. Vospriyatie kak veroyatnostnўi protsess. [Perception as a probabilistic process.] *Vop. Psikhol.*, 1961, 7(5), 47–62.

BZHALAVA, I. T. Kontrastnaya illyuziya ili effekt posledeĭstviya figury. [Contrast illusion or figural aftereffect.] *Vop. Psikhol.*, 1962, 8(5), 57–69.

CHISTOVICH, L. A., ALYAKRINSKIĬ, V. V., & ABUL'YAN, V. A. Vremennўe zanderzhki pri povtorenii slўshimoĭ rechi. [Temporal delays in repetition of apprehended speech.] *Vop. Psikhol.*, 1960, 6(1), 114–120.

CHISTOVICH, L. A., KLASS, YU. A., & ALEKIN, R. O. O znachenii imitatsii dlya raspoznavaniya zvukovўkh posledovatel'nosteĭ. [On the significance of imitation for discrimination of sound sequences.] *Vop. Psikhol.*, 1961, 7(5), 173–182.

CHZHI-TSIN, V., ZINCHENKO, V. P., & RUZSKAYA, A. G. Sravnitel'nўi analiz osyazaniya i zreniya: Soobschenie VIII. Nekotorўe priemў sensornogo obucheniya. [Comparative analysis of touch and vision: Communication VIII. Some techniques of sensory training.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1961, 5(4), 73–76.

DYMERSKIĬ, V. YA. *O vospriyatii prostranstvenno-vremennўkh otnoshenni.* [On the perception of spatial-temporal relations.] (Authorized summary of a candidate's dissertation.) Moscow: Moscow State University, 1960.

FONAREV, A. M. Soglasovannost' v dvizheniyakh glaznўkh yablok u novorozhdennўkh deteĭ i voprosy vospriyatiya prostranstva. Soobschenie I. [Conjunction of eye movements in neonates and the question of depth perception. Communication I.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1959, 3(4), 85–88.

GARNER, W. R., & HAKE, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, 58, 446–459.

GAYDOS, H. F. Intersensory transfer in the discrimination of form. *Amer. J. Psychol.*, 1956, 69, 107–110.

GIPPENREITER, YU. B. Analiz sistemnogo stroeniya slukhovogo vospriyatiya: Soobschenie I. K metodike izmereniya raslichi-

tel'noĭ chuvstvitel'nosti. [Analysis of the systemic structure of auditory perception: Communication I. On the method of measurement of differential sensitivity.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1957, **1**(4).

IPPENREITER, YU. B. Analiz sistemnogo stroeniya slukhovogo vospriyatiya: Soobschenie II. Eksperimental'nyĭ analiz motornoĭ osnovnyŷ protsessa vospriyatiya vysoty zvuka. [Analysis of the systemic structure of auditory perception: Communication II. Experimental analysis of the motor basis of the process of perception of the pitch of sound.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1958, **2**(1), 47–50.

GIPPENREITER, YU. B. *O vospriyatii vŷsotŷ zvuka.* [On the perception of the pitch of sounds.] (Authorized summary of dissertation for the degree of candidate of pedagogical sciences.) Moscow: Academy of Pedagogical Sciences RSFSR, 1960.

HOLWAY, A. H., & BORING, E. G. Determinants of apparent visual size with distance variant. *Amer. J. Psychol.,* 1941, **54**, 21–37.

KALIKINSKIĬ, Yu. A. Issledovanie nekotorŷkh funktsii slukha u spetsialistov po tekhnicheskomu oslushivaniya. [A study of some functions of hearing in technicians who use sound indices.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1961, **5**(1), 89–92. (a)

KALIKINSKIĬ YU. A. Nekotorŷe osobennosti slukha spetsialistov po tekhnicheskomu oslushivaniu dvgateleĭ vnutrennego sgoraniya. [Some characteristics of hearing of specialists in technical inspection of internal combustion engines.] *Vop. Psikhol.,* 1961, **7**(5), 121–134. (b)

KHOMSKAYA, E. D. K probleme afferentsii dvizhenii glaz. [On the problem of afferentiation of eye movements.] *Vop. Psikhol.,* 1962, **8**(3), 73–84.

LAVRENT'EVA, T. V., & RUZSKAYA, A. G. Sravnitel'nyĭ analiz osyazaniya i zreniya: Soobschenie V. Odnovremennoe intersensornoe sopostavlenie formŷ v doshkol'nom vozraste. [Comparative analysis of touch and vision: Communication V. Simultaneous intersensory comparison of form at a preschool age.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1960, **4**(4), 73–76.

LENIN, V. I. *Materialism and emperio-criticism.* New York: International Publishers, 1927.

LEONT'EV, A. N. Ob istoricheskom podkhode v izuchenii psikhi cheloveka. [On the historical approach to the study of the mind of man.] In B. G. Anan'ev (Ed.), *Psikhologicheskaya nauka v SSSR.* Tom I. [Psychological Science in USSR. Vol. I.] Moscow: Publishing House of the Academy of Pedagogical Sciences RSFSR, 1959. (a)

LEONT'EV, A. N. O mekhanisme chustvennogo otrazheniya. [On the mechanism of sensory reflection.] *Vop. Psikhol.,* 1959, **5**(2), 19–41. (b)

LEONT'EV, A. N., & GIPPENREITER, YU. B. Analiz sistemnogo stroeniya vospriyatiya: Soobscheni VIII. Vliyanie rodnogo yazŷka na formirovanie slukha. [Analysis of the systemic structure of perception: Communication VIII. Influence of native language on the formation of hearing.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1959, **3**(2), 59–62.

LIAMINA, G. M. Razvitie ponimaniia rechi u deteĭ vtorogo goda zhizni. [Development of comprehension of speech by children during their second year of life.] *Vop. Psikhol.,* 1960, **6**(3), 106–121.

LURIYA, A. R., PROVIDINA-VINARSKAYA, E. N., & YARBUS, A. L. K voprosu o mekhanizmakh dvizhenii glaz v protsessa zritel'nogo vospriyatiya i ikh patologii. (Simul'tannaya agnoziya i opticheskaya ataksiya.) [On the question of the mechanisms of eye movements in the process of visual perception and their pathology.] *Vop. Psikhol.,* 1961, **7**(5), 159–172.

MARUSEVA, A. M. Ob izmereniyakh slukha u deteĭ pervŷkh let zhizni. [On measurements of hearing in children of the first year of life.] *Probl. fiziol. Akust.,* 1959, **4**, 123–132. (a)

MARUSEVA, A. M. O slukhovoĭ chuvstvitel'nosti chelovka v usloviyakh trenirovki. [On auditory sensitivity of man in training.] *Probl. fiziol. Akust.,* 1959, **4**, 100–106. (b)

MILERYAN, E. A., & TKACHENKO, V. G. O vliyanii uprazheniya na velichenii prostranstvennogo poroga taktil'nogo razlicheniya. [The influence of practice on the magnitude of spatial threshold in tactual discrimination.] *Vop. Psikhol.,* 1961, **7**(3), 61–69.

NATADZE, R. G. Ob odnom fatore konstantnosti vospriyatiya velichinŷ. [On one factor in size constancy.] *Vop. Psikhol.,* 1960, **6**(3), 24–32.

NATADZE, R. G. K probleme konstantnosti vospriyatiya. [On the problem of constancy of perception.] *Vop. Psikhol.,* 1961, **7**(4), 35–52.

OVCHINNIKOVA, O. V. Analiz sistemnogo stroeniya slukhovogo vospriyatiya: Soobschenie III. O vliyanii zagruzki golosouŷkh svyazok na otsenky vŷsotŷ pri zvukorazlichenii. [Analysis of the systemic structure of auditory perception: Communication III. On the influence of activity of the vocal chords on evaluation of pitch discrimination.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1958, **2**(1), 51–54.

OVCHINNIKOVA, O. V. Analiz sistemnogo stroeniya vospriyatiya: Soobschenie VI. O "sensornoĭ" trenirovke zvukovŷsotnogo slukha. [Analysis of the systemic structure of perception: Communication VI. On the "sensory" training of pitch perception.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1959, 3(1), 79–82. (a)

OVCHINNIKOVA, O. V. Analiz sistemnogo stroeniya vospriyatiya: Soobschenie VII. Trenirovki slukha po "motornoĭ" metodike. [Analysis of the systemic structure of perception: Communication VII. Training of hearing by the "motor" method.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1959, 3(2), 55–58. (b)

OVCHINNIKOVA, O. V. Analiz sistemnogo stroeniya vospriyatiya: Soobschenie IX. Opŷt zamescheniya motornogo zvena v sisteme zvukovŷsotnogo slukha. [Analysis of the systemic structure of perception: Communication IX. An experiment in replacing of the motor link in the system of pitch perception.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1960, 4(3), 79–86. (a)

OVCHINNIKOVA, O. V. Opŷt formirovaniya zvukovŷsotnogo slukha. [An experiment in the formation of pitch perception.] (Authorized summary of dissertation for the degree of candidate of pedagogical sciences.) Moscow: Academy of Pedagogical Sciences RSFSR, 1960. (b)

PRANGISHVILI, A. S., & KHODZHAVA, Z. I. (Ed.) *Eksperimental'nŷe issledovaniya po psikhologii ustanovki.* [Experimental investigations in the psychology of set.] Tbilisi: Publishing House of the Academy of Science, Georgian SSR, 1958.

SOKOLOV, E. N. Veroyatnostnaya model' vospriyatiya. [Probabilistic model of perception.] *Vop. Psikhol.,* 1960, 6(2), 61–73.

SOKOLOV, E. N., & MIKHALEVSKAYA, M. B. Izmerenie effektivnosti deĭstviya mnogokratno nanosimogo razdrazhitel'ya. [Measurement of the effectiveness of a repeatedly presented stimulus.] *Vop. Psikhol.,* 1961, 7(5), 183–188. (a)

SOKOLOV, E. N., & MIKHALEVSKAYA, M. B. K voprosu izmenenii svyazi mezhdu stimulom i reaktsii. [On the question of change of connection between stimulus and reaction.] *Vop. Psikhol.,* 1961, 7(1), 57–72. (b)

SOKOLOV, E. N., & MIKHALEVSKAYA, M. B. Metod testiruyuschego stimula. [The method of the probe stimulus.] *Vop. Psikhol.,* 1962, 8(1), 28–36.

TARAKANOV, V. B., & ZINCHENKO, V. P. Sravnitelnŷi analiz osyazaniya i zreniya: Soobschenie VI. Proizvol'noe i neproizvol'noe zapomenanie formŷ v doshkol'nom vozrasti. [Comparative analysis of touch and vision: Communication VI. Voluntary and involuntary memory of form in preschool children.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1960, 4(5), 49–52.

VEKKER, L. M., & LOPE, YU. P. K probleme postroeniya osyazatel'nogo obraza. [On the problem of the formation of tactual images.] *Vop. Psikhol.,* 1961, 7(5), 143–153.

WETTER, G. G. *Dialectical materialism.* New York: Praeger, 1958.

ZINCHENKO, V. P. Nekotorŷe osobennosti orientirovochnŷkh dvizhenii ruki i glaza i ikh rol' v formirovanii dvigatel'nŷkh navŷkov. [Some properties of orienting movements of the hands and eyes and their role in the formation of motor habits.] (Authorized summary of candidate's dissertation.) Moscow: Institute of Psychology, 1957.

ZINCHENKO, V. P. Sravnitel'nŷi analiz osyazaniya i zrenya: Soobschenie II. Osobenesti orientirovochno-issledovatel'skikh dvzhenii glaza u deteĭ doshkol'nogo vozrasta. [Comparative analysis of touch and vision: Communication II. Properties of orienting-investigatory eye movements in preschool children.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1960, 4(2), 53–60.

ZINCHENKO, V. P., CHZHI-TSIN, V., & TARAKANOV, V. V. Stanovlenie i razvitie depertstivnŷkh deĭstvii. [Formation and development of perceptive behavior.] *Vop. Psikhol.,* 1962, 8(3), 1–14.

ZINCHENKO, V. P., LOMOV, B. F., & RUZSKAYA, A. G. Sravnitel'nŷi analiz osyazaniya i zreniya: Soobschenie I. O tak nazivaem "simultanoe" vospriyatie. [Comparative analysis of touch and vision: Communication I. On so-called simultaneous perception.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1959, 3(5), 71–74.

ZINCHENKO, V. P., & RUZSKAYA, A. G. Sravnitel'nŷi analiz osyazaniya i zreniya: Soobschenie III. Zritel'no-gapticheskiĭ perenos v doshkol'nom vozraste. [Comparative analysis of touch and vision: Communication III. Visual-haptic transfer in preschool age.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1960, 4(3), 95–98. (a)

ZINCHENKO, V. P., & RUZSKAYA, A. G. Sravnitel'nŷi analiz osyazaniya i zreniya: Soobschenie VII. Nalichnŷe uroveni vospriyatiya formŷ u deteĭ doshkolnogo vozrasta. [Comparative analysis of touch and vision: Communication VII. The observable level of perception of form in children of preschool age.] *Dokl. Akad. Pedag. Nauk RSFSR,* 1960, 4(6), 85–88. (b)

(Received May 14, 1963)

Psychological Bulletin
1964, Vol. 62, No. 1, 36–55

A MULTIPLE-INDICATOR APPROACH TO ATTITUDE MEASUREMENT[1]

STUART W. COOK[2] AND CLAIRE SELLTIZ
New York University

Most experimental research on attitudes has used crude measuring instruments, relying on aspects of research design and analysis to overcome ambiguities of interpretation permitted by the measuring instrument. As a step toward more efficient selection of measuring instruments, this paper examines different types of instruments from the point of view of the kinds of evidence they provide as a basis for assessing attitudes and of the nature of the inferences involved. 5 classes of techniques are discussed—measures in which the material from which attitudes are inferred consists of: self-reports of beliefs, feelings, behavior, etc., toward an object or class of objects; observed overt behavior toward the object; reactions to or interpretation of partially structured material relevant to the object; performance on "objective" tasks where functioning may be influenced by disposition toward the object; and physiological reactions to the object.

At least since LaPiere's report (1934) of the discrepancy between the actual reception accorded him and a Chinese couple and the answers to a questionnaire about accepting Chinese as guests, investigators have been concerned with the fact that different procedures designed to assess the same attitudes have often led to quite different placements of the same individuals, and that observed behavior toward a social object (person, group, etc.) is frequently not what would have been predicted from a given instrument intended to measure attitude toward that object. There have been several types of reaction to such observed discrepancies. One has been to assume that there is a "true" attitude toward the object, which one or both measures have failed to gauge correctly. A second has been to assume that there are different "classes" of attitudes toward a given object—for example, "verbal attitudes" and "action attitudes"—which should not necessarily be expected to correspond. Another has been to equate attitude with behavior, using "attitude" simply as a descriptive term summarizing observed consistencies in behavior. Still another reaction has been to think of attitude as an underlying disposition which enters, along with other influences, into the determination of a variety of behaviors toward an object or class of objects, including statements of beliefs and feelings about the object and approach-avoidance actions with respect to it.

We prefer the latter position; first, because for us, as for others (e.g., Allport, 1954) the observation of regularities in social behavior seems to point to the operation of relatively stable under-

[1] This paper was prepared as part of a program of research on the measurement of social attitudes supported by grants from the National Science Foundation and the Air Force Office of Scientific Research. Many of our colleagues have contributed to our thinking about the problems discussed here. In particular, we have gained insights and clarified our ideas through discussions with Donald T. Campbell, Isidor Chein, Barbara S. Dohrenwend, John S. Harding, Marie Jahoda, Irwin Katz, and M. Brewster Smith. It is perhaps unnecessary to add that there is no implication that any of these colleagues completely share the point of view expressed here.

[2] Now at the University of Colorado.

lying dispositions toward classes of objects. Further, we believe that apparent inconsistencies in social behavior may often best be understood in terms of the operation of such stable underlying dispositions in shifting relation to other influences on behavior. Finally, if validly distinguished, a dispositional concept has, by its very nature, a wider range of situational relevance—including projectability into relatively novel situations—than a simple descriptive concept equating attitude with behavior in specified situations.

We assume that two classes of variables, in addition to an individual's attitudinal disposition toward a given object or class of objects, influence his behavior in situations involving the object or symbols of the object (including the behavior constituting his responses to instruments designed to measure attitude toward the object): (*a*) *other characteristics of the individual,* including his dispositions toward other objects represented in the situation, values he holds that are engaged by the situation, his motivational state, his expressive style, and so on; (*b*) *other characteristics of the situation,* including its prescriptions as to appropriate behavior, the expectations of others in the situation with respect to the individual's behavior, the possible or probable consequences of various acts on his part, and so on.

In this view, an attitude cannot be measured directly, but must always be inferred from behavior—whether the behavior be language in which the individual reports his feelings about the attitude-object, performance of a task involving material related to the object (e.g., recall of statements which take a position with respect to the object), or actions toward a representative of the object-class (e.g., avoidance of such an individual). Lazarsfeld (1959) takes a similar position in his discussion of latent structure analysis. He points out that there is a probability relation between an indicator and the underlying trait of which it is taken as an indication; that is, a given trait does not invariably produce a given behavior. He stresses that, in consequence, some inconsistency will always be found between different measures of a hypothesized trait, and that the task of the investigator is to combine them into an "index" or "measurement" which represents the best inference that can be made from the manifold of empirical operations to the underlying characteristic they are assumed to reflect.

This orientation leads to emphasis on the need for a number of different measurement approaches to provide a basis for estimating the common underlying disposition, and to the expectation that data from these approaches will not be perfectly correlated. However, it seems to us that it should be possible to increase the correspondence among the indicators by careful analysis of other factors that are likely to affect response to a given measuring instrument and by efforts to reduce or control the influence of those factors. Ideally, the goal would be to develop one or more measures from which the effects of all probable response determinants other than attitude toward the relevant object would be removed. This goal, however, seems unlikely of achievement; therefore it seems to us important to work with a number of different measures, in each of which an effort is made to eliminate or control in some systematic way some identifiable influence on response other than the attitude in question. Since different influences will be controlled in different measures—and thus, conversely, different influences in addition to attitude will affect responses on the different measures—there will remain a lack of full correspondence among scores on the different measures.

Social scientists have long recognized that factors other than an individual's attitude toward an object may influence both his response to instruments designed to measure the attitude and his behavior toward the object in everyday life. Much recent work in the field of both personality and attitude measurement has been concerned with identifying the effects of such "extraneous" variables as the tendency to agree (or to disagree) with statements regardless of their content (e.g., Bass, 1955; Cronbach, 1946, 1950) or the wish to give a socially acceptable picture of oneself (e.g., Edwards, 1953, 1957; Taylor, 1961). Another interest has been in the development of indirect methods of attitude assessment (for a review of such methods, see Campbell, 1950). But attempts to develop indirect measures have, for the most part, been sporadic, and there has been little effort to examine systematically the relation of different indirect measures to each other or their relative susceptibility to such influences as agreeing response set or social norms.

Despite the general awareness of measurement problems, examination of reports of experimental research on attitudes shows the following picture: First, even investigators who hold very sophisticated theoretical positions about the nature and functions of attitudes and the conditions for attitude change commonly use only a single attitude measure—typically quite crude—in testing hypotheses derived from those theoretical positions. Second, most investigators are aware of the possibility that responses to these instruments may be influenced by factors other than the attitudes they are intended to measure. Third, efforts are made to guard against the intrusion of such factors or to rule out interpretations based on the possibility that they have been operative. These safeguards usually take one or more of the following forms: sampling (e.g., selection of groups of subjects believed to differ in susceptibility to the extraneous influences most likely to be operative in the measurement situation), experimental design (e.g., the introduction of control groups), internal analysis of the data (e.g., considering how the responses of subgroups of subjects might be expected to differ if one determinant rather than another were operative).

We do not mean to minimize the importance of such procedures. In any given study they may quite convincingly rule out the possibility that responses have been influenced by factors other than subjects' attitudes toward the object in question. Nevertheless, it seems to us that effort directed toward improving measuring instruments might be at least equally useful.

An Examination of Different Types of Measuring Instruments in Terms of the Kinds of Evidence They Provide as a Basis for Inferences about Attitude

In most current research on attitudes, efforts directed specifically toward improving measuring techniques are limited to such matters as assuring anonymity, attempting to separate the measurement from the experimental sessions, varying the order of presentation of items or the context in which they are embedded. If we are to go beyond such limited steps, a more systematic analysis of the characteristics of measuring instruments is needed than is yet available. This paper is a first step toward such an analysis. Our purpose is not to present a detailed review of the different kinds of instruments that have been used to measure attitudes; this has been well done by others (Campbell, 1950; Deri, Dinnerstein, Harding, & Pepitone, 1948; Weschler & Bernberg, 1950). Rather, we propose to examine broad

classes of measurement techniques from the point of view of the kinds of evidence they provide and thus the nature of the inferences involved in estimating attitude. By "the nature of the inferences involved" we mean the grounds for believing that attitude toward the presumed object is a determinant of responses to the measuring instrument, and the bases for inferring the nature of the attitude from the characteristics of the responses (i.e., for considering a given response as indicative of a positive or a negative disposition toward the object).

We have found it useful to think in terms of five major groupings: (*a*) measures in which the material from which inferences are drawn consists of self-reports of beliefs, feelings, behavior, etc., toward an object or class of objects; (*b*) measures in which inferences are drawn from observed overt behavior toward the object; (*c*) measures in which inferences are drawn from the individual's reactions to or interpretations of partially structured material relevant to the object; (*d*) measures in which inferences are drawn from performance on objective tasks where functioning may be influenced by disposition toward the object; and (*e*) measures in which inferences are drawn from physiological reactions to the object. Not all of the measures discussed have been used as attitude tests in the formal sense, but for each of them there is reason to believe that attitude may be an important determinant of response and thus that the technique could serve as a basis for inferences about attitude.

In assessing the adequacy of an instrument as an indicator of attitude, consideration of its susceptibility to other influences is as important as consideration of the grounds for believing that underlying disposition toward the object is a determinant of response. In examining measuring instruments from the point of view of the possible influence of factors other than attitude, we shall consider two major aspects: (*a*) the probability that overt responses may deviate from "private" responses—that is, the ease with which an individual can alter his responses to present a certain picture of himself; (*b*) the probability that private responses may be influenced by determinants other than attitude, in the absence of any attempt to distort responses.

Possibilities of influence of private response by factors other than attitude are, of course, almost limitless; we shall discuss only those that seem most probable with respect to each type of instrument. Susceptibility of overt response to distortion—that is, the possibility of discrepancy between private and overt response—would seem to be a function of three characteristics of the instrument: the extent to which its purpose is apparent, the extent to which the implications of specific responses are clear, and the extent to which responses are subject to conscious control.

In discussing the susceptibility of measures to distortion of responses and techniques developed to lessen the probability of distortion, we assume that with respect to many attitudes the settings in which tests are usually administered tend to exert pressures in a constant direction. It seems reasonable to suppose that most respondents, presented with tests in an academic setting or under the auspices of some other "respectable" organization, will assume that the responses which will place them in the most favorable light are those which represent them as well adjusted, unprejudiced, rational, open minded, and democratic. Moreover, since these are ideal norms at least in much of the American middle class, the pressures specific to the test situation are likely to coincide with inner pressures toward maintaining an image acceptable to the

self as well as to others. By "controversial social attitudes" we mean attitudes with respect to which such norms are operative. Some of our discussion, and especially some of our examples, concern techniques for making it easier for the individual to reveal himself as not well adjusted, not unprejudiced, etc., or for making it harder for him to portray himself, falsely, as well adjusted, unprejudiced, etc. While some assumption as to the probable direction of pressures operating in the situation is necessary for the concrete details of certain techniques, the principles involved do not hinge on the specific direction of pressures; given testing situations in which there is reason to believe that the pressures are predominantly in a different direction, the techniques can be modified accordingly. And many of the techniques require no assumption about the probable direction of pressures, being designed to reduce the effects of extraneous influences in any direction.

Measures in Which Inferences Are Drawn from Self-Reports of Beliefs, Feelings, Behaviors, etc.

By far the most frequently used method of securing material from which to make inferences about an attitude is to ask an individual to reveal—either in his own words or through acceptance or rejection of standardized items—his beliefs about the attitudinal object, how he feels toward it, how he behaves or would behave toward it, how he believes it should be treated.

The basis for inference is clear: it is axiomatic in all definitions that an individual's attitude toward an object is indicated by his beliefs, feelings, and action orientation toward it. The nature of the inference is also clear: it is assumed that the relationship between attitude and expression is a direct one and that the attitude corresponds to the manifest, common-sense implications of the stated belief or feeling. For example, a stated belief that the object has characteristics usually considered desirable is taken as reflecting a favorable disposition toward it, and a stated belief that it has characteristics usually considered undesirable is taken as reflecting an unfavorable disposition. Similarly, a report that the person avoids contact with the object is taken as indicating an unfavorable disposition toward it, while a report that he does or would willingly enter into contact with it is taken as indicating a favorable disposition.

In some definitions, attitude is considered identical with, or simply a summary of, beliefs, feelings, behavior, etc., toward the object; thus no problem of inference arises. However, in such definitions some criteria must be adopted for choosing which behavior constitutes the population of "attitudinal responses" to be sampled. The choice of such criteria would, we believe, depend upon an analysis essentially similar to our consideration of "extraneous influences" in the remainder of this paper.

Self-report measures have a number of characteristics that make them susceptible to distortion of overt responses. The purpose of the instrument is obvious to the respondent; the implications of his answers are apparent to him; he can consciously control his responses. Thus a person who wishes to give a certain picture of himself—whether in order to impress the tester favorably, to preserve his own self-image, or for some other reason—can rather easily do so. This difficulty has long been recognized, and in recent years it has been extensively investigated under the rubric of "social desirability." A number of techniques have been devised to make the purpose of the instrument or the implications of the responses less apparent; to make it easier to give answers that may be considered undesirable; and to make it harder to give, falsely, answers that

may be considered desirable. Some of these techniques are focused primarily on reducing the likelihood that responses will be distorted in an attempt to meet the investigator's expectations or to please him; others are addressed to reducing the influence on responses of a desire to maintain a certain self-image as well as that of a desire to please or impress the investigator.

One of the simplest approaches to making the purpose of the instrument less apparent is the inclusion of items not relevant to the attitudinal object in which the investigator is interested. A variation of this approach is to include in each of the items a number of aspects in addition to that in which the investigator is interested; for example, if the investigator is interested in attitudes toward one or more racial groups, each item may refer to a hypothetical person characterized not only in terms of race but of age, sex, religion, occupation, etc. Approaches of either sort serve only to make the purpose of the test less obvious. They do not completely conceal or disguise it, nor can they do so within the format of self-report measures, which by definition call for the individual's own account of his reactions to the attitudinal object.

Among the simplest, and most frequently used, approaches to making it easier to give answers that may be considered undesirable are assurances of anonymity, statements to the effect that "there are no right or wrong answers" or that "people differ in their views on these things," emphasis on the importance of honest answers in order to contribute to scientific knowledge or some other presumably desirable outcome, efforts to build up rapport between questioner and respondent and to create the impression that the questioner will not disapprove of whatever views may be expressed.

Other approaches are built into the instrument itself: including items to which an unfavorable reply is likely to be considered acceptable (e.g., "Would you be willing to have a ditch digger as U. S. Congressman from your district?" —Westie, 1952, 1953), in order to break down a possible set to give uniformly favorable replies; including in the statement of a view that may be considered undesirable a qualification or a justification of it (e.g., "It is best that Jews should have their own fraternities and sororities, since they have their own particular interests and activities which they can best engage in together, just as Christians get along best in all-Christian fraternities"—Adorno, Frenkel-Brunswik, Levenson, & Sanford, 1950); wording questions in such a way that they assume the respondent holds certain views or has engaged in certain kinds of behavior (e.g., "When did you first . . . ?"—Kinsey, Pomeroy, & Martin, 1948).

Other approaches are designed to make it difficult to give, falsely, what may be considered a desirable answer In the measurement of personality, a major effort in this direction has been the use of forced-choice tests, where the respondent is asked to indicate which of two statements, matched in terms of social desirability but differing in their implications with respect to traits or needs, is closer to his own views or more descriptive of his own behavior. This approach has not been extensively used in the measurement of attitudes.

In addition to their susceptibility to conscious distortion in order to give the picture the individual wishes to present of himself, responses to self-report measures may be influenced by another set of characteristics presumably unrelated to attitude toward the object in question—characteristics frequently labeled "response set" or "expressive style." It has long been noted that some individuals have a consistent tendency to agree (or to disagree) with items presented to them, regardless of their content; or

to select, with more than chance frequency, the alternative which appears in a given position; or to give extreme (or moderate) answers.

A number of techniques have been devised to reduce the effects of such tendencies on scores that are to be taken as indicative of attitudes. Perhaps the simplest and the most common approach to the problem of influence by a tendency to agree (or to disagree) is to vary the wording of items in such a way that for approximately half of them agreement represents a favorable response to the attitudinal object, and for half an unfavorable response. Other approaches to this problem involve setting up the instrument in such a way that responses do not take the form of expressing agreement or disagreement with one statement at a time. The instrument may consist of pairs of statements representing roughly opposed points of view on a given issue, both statements being worded positively or both worded negatively; the subject is asked to indicate which is nearer his own position, or to indicate his position on a scale running between the two statements. The following pair of items from an unpublished scale of attitudes toward freedom of speech, developed by students of Donald T. Campbell at Northwestern University, illustrate this approach:

A. Fascists and Communists are entitled to preach their beliefs in this country.
B. Only those who are in agreement with this country's philosophy of government are entitled to preach their beliefs.

In other instruments, the problem, at least in its obvious form, is avoided by using items that call for free response—open-ended questions, sentence stubs to be completed with the individual's own responses, etc.

An approach to correcting for the effects of a tendency to give extreme answers, or moderate answers, consists in providing matched pairs of items, one referring to the attitudinal object, the other referring to some control object, and scoring in terms of the discrepancy between the two responses. For example, if respondents are asked only, "Would you be willing to have a Negro bookkeeper live in the same apartment building you live in?" and are provided with a 5-point response scale, it is impossible to determine whether respondents who answer "very willing" differ from those who answer simply "willing" in attitude, in response style, or in both. Providing a parallel item with respect to a white bookkeeper and scoring on the basis of discrepancy between an individual's responses to the Negro and the white removes the effects of response style from the score (Westie, 1953).

Susceptibility of self-report measures to the two kinds of influences discussed so far—desire to present a certain picture of oneself, and response sets unrelated to the content of items—clearly leads to the possibility of distortion of responses in the obvious sense of lack of correspondence between the overt responses and the individual's private beliefs, feelings, policy views, etc. Still other factors, however, may influence his private beliefs and feelings as well as his overt responses. While private beliefs, feelings, and action orientations with respect to an object are by definition at least partially determined by the individual's attitude toward the object, they may be influenced by other factors as well—for example, by the availability of information, or by other values the individual holds. Thus, a person who has an essentially devaluing attitude toward Negroes may nevertheless have learned and state as his belief that there is no difference in the chemical composition of the blood of Negroes and whites; on the other hand, a person whose disposition toward Ne-

groes is not devaluing may know and state as his belief that the average scholastic achievement of Negroes in the United States is lower than that of whites. A person with a devaluing attitude toward Negroes may nevertheless believe that they should not be deprived of the right to vote, because he sees this right as an essential ingredient of democracy; a person whose attitude toward Negroes is not devaluing may be opposed to laws forbidding discrimination in the sale and rental of housing because he places great store on the right of an owner to do with his property as he sees fit.

To the extent that such other influences affect different items differently, or affect only certain items, this problem has been attacked by examining responses for consistency, eliminating items which show low agreement with total scores, or eliminating those to which responses do not fall on a unidimensional scale.

A given technique may help to reduce or correct for extraneous influence from more than one source. For example, scoring in terms of discrepancy between responses to items concerning the attitudinal object and comparable items about a control object may provide a correction for the effects of other values or meanings engaged by the items as well as for response sets. Asking the respondent to choose which of two statements is closer to his views may help to eliminate the influence both of response set and of concern with the acceptability of responses, if the alternatives provided are equivalent in both respects.

Not only may a given technique serve more than one function; a given instrument may embody a number of techniques designed to reduce the influence of extraneous factors. For example, in Westie's (1953) Summated Differences Test, the subject is presented with hypothetical persons of specified race (Negro or white) and occupation (eight occupations, ranging from ditch digger to banker, plus "the average man"), and asked to indicate, on 5-point scales, his willingness to accept each of these 18 hypothetical persons in each of 24 relationships—a total of 432 items. Some of the items are such that a negative answer is likely to be considered acceptable by most people (e.g., unwillingness to vote for a machine operator, whether white or Negro, as President of the United States is not likely to be seen as an expression of "prejudice"), thus presumably breaking down a possible tendency to give uniformly favorable answers whether through an acquiescent response set or through a desire to give a picture of oneself as unprejudiced. The large number of items, and the format of the questionnaire, make it extremely unlikely that the subject can remember or check his response to a given item with respect to one racial group when he is answering the comparable item concerning the other group. Scoring on the basis of discrepancy between parallel items referring to whites and Negroes takes account both of possible response sets and of the influence of the specified occupation and the specified situation. Thus, this instrument adds to the basic social distance questionnaire a number of techniques designed to make the focus of the investigator's interest less apparent, to make it easier to give answers that might be considered undesirable, to correct for possible response sets, and to some extent to take account of other values or meanings that may affect responses.

Measures in Which Inferences Are Drawn from Observation of Overt Behavior

Many investigators have pointed out the desirability of using measures in which overt behavior toward members

of a class of objects would serve as a basis for inferences about attitude toward the object-class. As with self-report measures, the basis for inference is clear; all definitions of attitude specify that behavior can be taken as an indicator of attitude. And, as in the case of self-report measures, the usual assumption is that there is a simple correspondence between the nature of the behavior and the nature of the underlying attitude; for example, that friendly behavior toward a member of a given class of objects indicates a favorable attitude toward the object-class.

There has been much less extensive development of measures of this sort than of self-report measures. Situations capable of eliciting behavior toward an attitudinal object are more difficult to devise and to standardize, and more time consuming and costly to administer, than self-report measures. Although some measures of this type have been devised, they have not been widely enough used to provide much evidence as to their specific strengths and weaknesses nor to stimulate efforts to correct for shortcomings. However, analysis of their characteristics can provide estimates as to their probable susceptibility to influences other than attitude and possibilities of reducing such susceptibility.

Attempts to develop behavioral measures have followed three general lines. One consists in presenting subjects with standardized situations that they are led to believe are unstaged, in which they believe that their behavior will have consequences, and in which the attitudinal object is represented in some way other than by the actual presence of a member of the object-class. For example, subjects may be asked to sign a petition on behalf of an instructor about to be discharged for membership in the Communist party, to contribute money for the improvement of conditions for migratory workers, to indicate whether they would be willing to have a Negro roommate. DeFleur and Westie (1958) have attempted to develop a measure of this sort which is appropriate for use in many different testing situations. In their procedure, as part of a larger program of research, white subjects viewed a number of colored photographic slides showing a young Negro man and a young white woman, or a young white man and a young Negro woman, in a social setting; subjects described the pictures and answered specific questions about them. At the close of an interview following this session the measurement procedure being discussed here was introduced. DeFleur and Westie describe the procedure as follows: The subject was told that another set of such slides was needed for further research, was asked if he (or she) would be willing to be photographed with a Negro of the opposite sex, and then was given "a standard photograph release agreement," containing a variety of uses to which such a photograph would be put, ranging from laboratory experiments where it would be seen only by professional sociologists, to a nationwide publicity campaign advocating racial integration. The subject was asked to sign his name to each use of the photograph which he would permit. These investigators report that subjects "uniformly perceived the behavioral situation posed for them as a highly realistic request."

Such devices differ from self-report measures with similar content in that, in the behavioral measures, the subject either actually carries out the behavior (signs a petition, makes a contribution, etc.) or is led to believe that his agreement to do so will lead to real-life consequences (being asked to pose for a photograph to be put to specified uses, being assigned a Negro roommate, etc.).

Another approach is to present the subject with an admittedly staged situa-

tion and ask him to play a role—perhaps to behave as he would in such a situation in real life, perhaps to take the part of someone else or to act in some specified way. Stanton and Litwak (1955) presented actual and potential foster parents with situations of interpersonal stress in which they were instructed to behave in a given way (defined as not manifesting specified undesirable or neurotic kinds of behavior); for example, in one scene the subject was instructed that he was to play the role of a married man, having dinner with his parents; the investigator, playing the role of the man's father, treated his son like a child, criticized his wife, and put him in the wrong. These investigators found that ratings based on a half-hour's role playing were better predictors of subjects' behavior as foster parents (as rated by case workers who had sustained contact with them) than were ratings based on 12 hours of intensive interviewing by a trained social worker. Stanton, Back, and Litwak (1956) reported that a role-playing approach was successful in discovering the limits of positive and negative feelings about public housing projects on the part of slum dwellers in Puerto Rico. These investigators have stressed the importance of designing the scene specifically to elicit responses relevant to the particular behavior or attitude in which the investigator is interested.

A third behavioral approach, used in the study of attitudes toward social groups, has been to ask for sociometric choices among individuals some of whom are members of the object group, preferably under circumstances that lead the participants to believe that such choices will have consequences in the form of subsequent assignment in some situation. Early applications of this technique to the study of intergroup attitudes were made in studies by Moreno (see 1943) and by Criswell (1937, 1939), in which patterns of choices by school children were analyzed in terms of the development of cleavage along racial lines. Subsequently, sociometric techniques have been used in research evaluating the effects of certain experiences on attitudes (e.g., Mann, 1959a; Mussen, 1950a, 1950b) and of the relations among different aspects of attitudes (e.g., Mann, 1959b).

There are differences among these three kinds of behavioral measures—situations appearing to the subject to be unstaged, role playing, and sociometric choice—in characteristics that affect the probability that overt responses will correspond to responses that would be shown if the individual were not concerned with presenting (to others or to himself) a certain picture of himself. Let us consider first the extent to which their purpose is apparent to the respondent. To the extent that the purportedly unstaged situations are accepted as genuine, the respondent will not see them as designed to get information about his attitudes; thus one possible source of pressure to give responses that are likely to be considered desirable is eliminated. Nevertheless, the implications of his behavior as revealing certain characteristics may be apparent to him; even if he accepts a question about his willingness to pose with a Negro or to have a Negro roommate as genuine, he may be aware that a positive answer will have the effect of presenting him as unprejudiced, a negative answer as prejudiced. Thus, even in the absence of awareness that he is being tested, an individual may be motivated to give a response that differs from his spontaneous private one, in order to present himself to the questioner as unprejudiced or to maintain his own image of himself as one who behaves in an unprejudiced way. The sociometric choice method would appear

to be similar in these respects, though it may perhaps be assumed that, in the absence of special influences calling attention to racial or ethnic group membership, the implications of the choices are less likely to be apparent. In the case of role playing, the extent to which the purpose of the situation and the implications of responses are clear presumably depends on the convincingness with which the situation can be presented as a measure of some other characteristic, such as acting ability.

All of these behavioral approaches have characteristics that may operate to make it easier to respond in ways that may be considered undesirable. In many situations it is possible to justify a negative response on neutral or acceptable grounds: one does not believe in signing petitions, or he does not like to have his picture taken, or he prefers Persons A and B to X and Y because they share his interest in music. Or, in the role-playing situation, his behavior is shaped not by his own reactions toward the attitudinal object but by interest in the dramatic requirements of the situation. (To the extent that these alternative explanations are real possibilities, however, they introduce other problems about interpretation of the behavior as an indicator of the attitude in which the investigator is interested.)

Some characteristics of the behavioral approaches may reduce the probability that the individual will modify his behavior in order to present an acceptable picture of himself. When responses are expected to have real-life consequences, the anticipation of such consequences may counterbalance the wish to make a good impression. In a social distance questionnaire, if one wishes to present himself (to the tester, or to himself, or both) as unprejudiced, there is little effective pressure against saying that one would be willing to work with a Negro, or to have a Negro roommate; but if the question is posed in a context where a positive reply is seen as leading to assignment of a Negro as a co-worker or a roommate, one must weigh his willingness to accept that consequence against his wish to appear unprejudiced. In role playing, the pressure for quick response to unanticipated stimulus situations probably operates to lessen conscious control of behavior in order to produce a desired impression. Faced with the necessity of doing or saying something to keep the situation going, the individual may not have time to consider the impression he is making; to the extent that this is so, this approach may be thought of as reducing the individual's conscious selection of his response.

Thus behavioral measures seem to be less susceptible than simple self-report measures to distortion of response in the interest of presenting a certain picture of the self. But they are at least as susceptible as self-report measures to the effects of other extraneous influences. It has sometimes been suggested that the model of behavioral measures would be apparently unstaged situations in which a member of the object-class is present. But it is clear that behavior in everyday life situations (which this model seeks to approximate) is not determined exclusively by attitude toward the presumed attitudinal object. In the case of behavior toward minority groups, for example, social custom is a major determinant; in communities with segregated transportation systems, almost all white people—regardless of their attitudes toward Negroes or toward segregation—sit in the white section, whereas in communities with unsegregated transportation systems, very few white people—regardless of their attitudes—refuse to sit next to Negroes. Other values may override attitudes toward the presumed object; an individual who feels physical revulsion at

the experience of eating with Negroes may nevertheless do so because he has come to believe that the ideals of democracy, or religious principles of brotherhood, or the position of the United States in the eyes of the world, require that all men be treated as equals. Finally, other characteristics of the object individuals may predominate over their ethnic identification in determining response to them. Thus, LaPiere (1934) concluded that the factors which most influenced the behavior of hotel and restaurant personnel to the Chinese couple with whom he was traveling "had nothing to do with race"; rather, it was the quality and condition of their clothing, the appearance of their baggage, their cleanliness and neatness, and above all, their self-confident and pleasant manner, that determined reactions. Observations such as this suggest that, to the extent that one is interested in tapping generalized dispositions toward a given group rather than in predicting behavior in specific situations, behavioral measures that call for response to a symbolic representation of the group may be less subject to influence by extraneous factors than measures that call for response to members of the group who are physically present.

Campbell (1961) has suggested an approach to the use of behavior measures which is based on the premise that different situations have different thresholds for the manifestation of hostile, avoidant, or discriminatory behavior. He suggests that, in order to secure evidence about an individual's attitude, it is necessary to place him in a number of situations with differing thresholds—ranging, for example, from eating with a Negro at a business men's luncheon club (assumed to be a situation with a low threshold for nondiscriminatory behavior—in other words, one in which it is easy to behave in an unprejudiced way) to renting one's house to a Negro (assumed to have a high threshold for nondiscriminatory behavior). The lowest-threshold situation in which an individual exhibits discriminatory behavior would indicate his position on a scale of attitude with respect to the group in question. Such a procedure would be effective in taking account of pressures that are constant for all, or most, individuals; it would not, it seems to us, rule out the effects of differences in the strength for different individuals of such influences as concern with social approval, other values seen as relevant to the situation, etc.

Measures in Which Inferences Are Drawn from the Individual's Reaction to or Interpretation of Partially Structured Stimuli

The characteristic common to techniques in this category is that, while there may be no attempt to disguise the reference to the attitudinal object, the subject is not asked to state his own reactions directly; he is ostensibly describing a scene, a character, or the behavior of a third person. He may be presented with a photograph of a member of the object-class (usually a person of a given social group) and asked to describe his characteristics; or he may be presented with a scene in which members of the object-class are present and asked to describe it, to tell a story about it, to predict the behavior of one of the characters, etc. The stimulus material may be verbal rather than pictorial; for example, the subject may be asked to complete sentence stubs referring to a hypothetical third person.

The bases for inferences about attitudes are those common to all projective tests: assumptions that perception of stimuli that are not clearly structured is influenced by the perceiver's own needs and dispositions; that, asked to provide an explanation or interpretation for which the stimulus presented gives

no clear clue, the subject must draw on his own experience or his own dispositions or his own definitions of what would be probable or appropriate; that, asked to attribute behavior to others, especially under speed conditions, the most readily accessible source of hypotheses is the individual's own response disposition. As in self-report and behavioral tests, the usual assumption is that the expressed response corresponds directly to the individual's attitude; for example, that attribution of desirable characteristics to a member of a given group represents a favorable attitude toward that group, that interpretation of a scene as one in which there is hostility toward a member of a given group represents a hostile attitude toward the group, that attribution of a positive (or a negative) response to a hypothetical third person with respect to a given object reflects a positive (or a negative) disposition toward the object in question.

A major reason for the development of such techniques is the assumption that, by disguising the purpose of the instrument and the implications of responses, they lessen the probability of distortion of responses in the interest of presenting a certain picture of the self. They are presented to the respondent not as measures of attitudes but as tests of imagination, verbal fluency, ability to judge character, social sensitivity, or some such characteristic. To the extent that the respondent accepts these explanations, he presumably is unaware not only of the purpose of the test but of the implications of his responses as revealing his own attitudes. Even if the subject does realize that he is expressing his own attitude, it is assumed that it may be easier to express views that may be considered undesirable if one does not explicitly acknowledge them as his own. In some instances the questions asked are nonevaluative, so that the implications of one or another response are quite unlikely to be apparent to the respondent; for example, "What is the [nonexistent] colored man in the corner doing? [Horowitz & Horowitz, 1938]."

Questions have been raised, however, about the validity of the assumption that responses, even though spontaneous and undistorted, reflect the individual's own attitude toward the object. While it seems clearly established that an individual's response may reflect his own disposition, it is not certain that it necessarily does so. Given a scene in which the roles of Negro and white are ambiguous, an individual who describes the Negro as being in a menial position may be reflecting his own devaluing disposition toward Negroes; on the other hand, he may simply be reporting the arrangement most commonly observed in our culture. Similarly, the responses he attributes to a hypothetical third person may be based either on his own response disposition or on his estimate of how most people would react in such a situation.

Attempts to secure evidence as to whether responses to instruments of this type do in fact reflect the individual's own attitudes have followed two lines: examination of the correspondence between estimates of attitude based on these measures and estimates based on other measures (usually of the self-report type); and examination of data secured from instruments of this sort in the light of predictions about patterns of results.

Several studies have found significant correspondence between results of measures of this type and scores on self-report measures. Proshansky (1943) found high correlations between scores based on a standard self-report scale for measuring attitude toward organized labor and scores based on descriptions of briefly-exposed ambiguous pictures of relevant social situations. Riddleberger and Motz (1957) found that subjects who scored high and those who scored low on a self-report measure of attitude

toward Negroes differed in their explanations of how the people in a pictured interracial group had met. Sommer (1954), using a modified form of Brown's (1947) adaptation of the Rosenzweig Picture-Frustration Test, was able to identify with considerable success not only individuals who scored high and those who scored low on a self-report scale of attitude toward Negroes but a subgroup who had been instructed to respond to the Picture-Frustration Test as if they were unprejudiced, even though their self-report scores were unfavorable.

However, in view of the assumption that an important characteristic of tests of this type is their relative lack of susceptibility, as compared with self-report measures, to efforts to present a certain picture of the self, correspondence with scores based on self-report measures is a dubious criterion. Getzels (1951), recognizing this fact, approached the problem by predicting conditions under which speeded completions of third-person sentence stubs would differ from completions, by the same respondents, of the same sentence stubs presented in the first person. He made two predictions: (*a*) that first- and third-person responses would differ on items subject to strong social norms not fully internalized by all members of the group and would not differ on items not subject to such norms; and (*b*) that in the case of the former items, more socially acceptable answers would be given on the first-person form than on the third-person form. Both predictions were strongly supported. Getzels recognized the possibility that responses to the third-person form might be based on estimates of how most people would respond rather than on the subjects' own response dispositions. Accordingly, he asked the subjects to estimate how most people would respond to the items about Negroes, and found no difference between the average estimates made by those whose third-person responses had been favorable and those whose third-person responses had been unfavorable.

A number of techniques involving perception—in a more literal sense—of ambiguous or unstructured material may be considered in this category. For example, a number of psychologists have been investigating the possible relation of attitudes to perception of stimuli presented under stereoscopic conditions of binocular rivalry. Bagby (1957), presenting pairs of cards differing in cultural content (e.g., a bullfighter and a baseball player) to subjects from Mexico and the United States, found that Mexicans tended to see the card with Mexican content, North Americans those with content familiar in the United States. Pettigrew, Allport, and Barnett, (1958), presenting to residents of South Africa pairs of pictures of individuals from different racial groups, found that Afrikaners deviated most consistently from other groups in their responses, overusing the "European" and "African" categories, underusing "Colored" or "Indian."

A study by Bray (1950) made use of unstructured visual material in a different way. Taking off from Sherif's (1935) finding that estimates of movement in the autokinetic phenomenon are markedly influenced by the estimates given by others, Bray investigated the effects of estimates by confederates who were identified as members of minority groups. He had the hypothesis that the extent and direction of such effects would be influenced by the subject's attitude toward the minority group. Here the unstructured perceptual material did not refer to the attitudinal object, but simply provided an opportunity for expressing indirectly a response to the attitudinal object—the physically present, minority-group member.

Again, there are problems about the nature of the inferences that can be drawn. Bray, for example, did not find

the direct relationship he had predicted between attitude toward the minority group (as measured by self-report scales) and responses to the minority-group members' estimates. In the case of binocular rivalry, in what way, if at all, does attitude influence perception? Does one see the picture with the most familiar content? Does one see the member of the racial group toward which he is most favorable, or the one toward which he is most hostile, or of which he is most afraid?

Questions such as these point both to the need for further research on the usefulness of these techniques as measures of attitude and to potentially fruitful lines of investigation of the relation between attitudes and response to various kinds of materials under various conditions.

Measures in Which Inferences Are Drawn from Performance of "Objective" Tasks

Approaches in this category present the respondent with specific tasks to be performed; they are presented as tests of information or ability, or simply as jobs that need to be done. The assumption common to all of them is that performance may be influenced by attitude, and that a systematic bias in performance reflects the influence of attitude.

For example, the subject may be asked to memorize material, some of which is favorable to the attitudinal object, some unfavorable, perhaps some neutral or irrelevant. The assumption is that material congenial with the subject's own position will be learned more quickly and remembered longer. Some empirical support is available for this assumption; for example, in a study by Levine and Murphy (1943), using material about the Soviet Union, and one by Jones and Kohler (1958) using statements about segregation. Or the subject is given a test of "information," in which at least some of the items referring to the attitudinal object either have no correct answers or are so unfamiliar that it can be assumed that few if any respondents will know the correct answers; alternative responses believed (by the investigator) to indicate relatively favorable or relatively unfavorable dispositions toward the object are provided. The assumption here is that, when forced to make a guess on ostensibly factual questions where he has no objective basis for an answer, the subject is likely to choose the alternative most consistent with his own attitudinal disposition. This assumption, too, is supported by some empirical evidence; for example, studies by Hammond (1948) and Weschler (1950) of attitudes toward labor and toward Russia, and by Rankin and Campbell (1955) of attitude toward Negroes. Or the task may be a test of "reasoning," in which syllogisms or other logical forms are presented, and the subject is asked to indicate which of a number of conclusions can appropriately be drawn. Items referring to the attitudinal object are paralleled by similar items with neutral or abstract content; scoring is on the basis of the number and direction of errors on the attitudinally relevant items as compared with the control items. The assumption is that reasoning may be swayed by attitudinal disposition, and thus that errors on the attitudinally relevant items reflect the individual's own position, if the parallel neutral items have been answered correctly. Watson (1925), Morgan (1945), and Thistlethwaite (1950), among others, have developed instruments of this type. Thistlethwaite found a significant difference between Northern and Southern college students in frequency of errors on items dealing with Negroes (as compared with errors on the neutral items), and no corresponding difference on items dealing with Jews, women, or patriotism.

Other measures place the emphasis on the material being judged or on the

outcome to be achieved rather than on the ability involved in achieving it. For example, the subject is asked to sort items about the attitudinal object in terms of their position on a scale of favorableness-unfavorableness, ostensibly in order to help in the construction of a Thurstone-type scale. The assumption here is that the rater's own attitude toward the object—especially if it is extreme—influences his judgments of the favorableness of statements about the object. Despite the earlier belief that ratings of items for Thurstone scales are not affected by the raters' own attitudes, a number of recent studies (e.g., Hovland & Sherif, 1952) have found such effects.

It seems reasonable to suppose that most subjects accept these tasks at face value; presumably only someone with rather sophisticated knowledge of research techniques in the social sciences would be aware of their attitudinal implications. Thus it seems reasonable to suppose that they may be relatively impervious to distortion in the interest of presenting a desired picture of the self.

Again, however, there are questions about the nature of the inferences to be drawn. If a subject shows marked and consistent bias, it seems reasonable to infer that he has an attitude toward the object strong enough to affect his performance. If he does not show consistent bias, however, are we to infer that his attitude is not strong, or not consistent? In other words, how sensitive are such measures? Is it possible that individuals with equivalent attitudes differ in the extent to which their performance on such tasks is influenced by those attitudes?

Another problem has to do with the direction in which attitude influences the response, and, conversely, with the nature of the inference to be drawn from a given response. Responses may reflect either wishes or fears; a member of the Communist party may overestimate the number of Communists in the United States, but so may a member of the John Birch Society. A person who underestimates the number of Negro doctors in the United States may do so on the basis of his feeling that Negroes do not have the ability to become doctors, or he may do so on the basis of his belief that opportunities for Negroes to obtain medical training are limited.

Judgments of the favorableness or unfavorableness of statements are subject to a similar problem of interpretation. Hovland and Sherif (1952), working with items about Negroes, found that ratings by Negro subjects and by white subjects who actively supported desegregation differed from ratings by "average" and by anti-Negro white subjects. However, other investigators (e.g., Manis, 1960; Weiss, 1959), working with statements about different attitudinal objects, found that subjects with extreme attitudes—whether favorable or unfavorable—showed similar patterns of ratings, which differed from those made by subjects with moderate attitudes.

As with the preceding category, these problems of interpretation point to the need for caution in inferring the attitude of a given individual from a single test of this sort, but they seem to point also to the probable usefulness of further empirical investigation of the relation of scores based on such measures to those based on tests providing other grounds for inference.

Another group of measures presented as objective tasks or tests of ability focus on the extent to which the attitudinal object figures prominently in the subject's organization of his environment, that is, its salience for him. The kinds of data appropriate for inference about the salience of an attitudinal object differ in some respects from the kinds appropriate for inference about the nature or direction of the attitude. Measures of salience have been devel-

oped primarily with respect to attitudes toward social groups. They are of two types: techniques for assessing the tendency to classify individuals in terms of group membership, and techniques for assessing the tendency to subordinate individual differences to group identification.

One technique for assessing the tendency to classify individuals in terms of group membership, originated by Horowitz and Horowitz (1938), may be presented as a test of concept formation. It consists in presenting to a subject sets of photographs of individuals differing in race, sex, age, and socioeconomic status and asking him to select those which "belong together." For example, one set may contain photographs of three white boys, one white girl, and one Negro boy. If the subject replies that the white girl does not belong, this is taken to mean that for him sex is a more important basis for classification than race; if he replies that the Negro boy does not belong, the inference is that race is a more important category for him than sex.

Another technique for assessing the tendency to classify individuals in terms of group membership, presented as a test of memory, involves the clustering, in recall, of verbal symbols for which alternative classificatory principles are available. This technique rests on the finding from studies of verbal behavior that when words drawn from various categories are presented in random order, subjects tend to recall them in clusters, with several words representing a given category being recalled together even though they were not next to each other in the list presented. In studying the salience of race as a basis for classification, a subject would be presented, in random order, with names of people from several different occupational categories—for example, baseball players, musicians, political figures, actors, one name in each category being that of a Negro. The extent to which names of Negroes are grouped together in recall would provide the basis for inference as to the salience of race as a basis for classifying individuals.

A measure of the tendency to subordinate individual differences to group identification, originated by Horowitz and Horowitz (1938), consists in showing the subject a number of photographs of individuals of different ethnic groups and then asking him to identify, from a larger number of photographs, those he has already seen. The task is presented as one involving perception and/or memory. Scoring is in terms of the proportion of correct responses to individuals of a given social group as compared with the proportion of correct responses with respect to individuals of other groups. The inference here is that accuracy in identifying whether or not pictures of specific individuals of a given social group have previously been seen is decreased by the tendency to subordinate individual differences to group identification.

Seeleman (1940-41), using pictures of whites and Negroes, found a high correlation between scores on this measure and scores on a self-report questionnaire designed to measure attitude toward Negroes, with the less-favorable subjects less accurate in identifying whether the Negro pictures had previously been exposed. The question whether there is, in general, a correlation between salience of an attitudinal object and favorableness of disposition toward it is an interesting problem for empirical investigation.

Measures in Which Inferences Are Drawn from Physiological Reactions to the Attitudinal Object or Representations of It

At the opposite extreme from measures relying on a subject's verbal report

of his beliefs, feelings, etc., are those relying on physiological responses not subject to conscious control. These may be measures of a subject's reaction—for example, galvanic skin response (GSR), vascular constriction—to the presence of a member of the object group or to pictorial representations of situations involving members of the object group. For example, Rankin and Campbell (1955) compared GSRs obtained when the experimenter was a Negro with those obtained when the experimenter was white; Westie and DeFleur (1959) recorded GSR, vascular constriction of finger, amplitude and duration of heartbeat, and duration of heart cycle, while the subjects were viewing pictures of whites and Negroes in social situations. Hess and Polt (1960) have photographed pupillary constriction in response to unpleasant stimuli and pupillary dilation in response to pleasant stimuli.

Or the measures may involve responses, such as salivation, blinking, vascular constriction, that have been conditioned to a verbal stimulus, and, by a process of semantic generalization, appear in response to words or concepts that are similar in meaning to the original stimulus. For example, Volkova (1953) has reported a series of experiments in Russia in which subjects were conditioned to salivate in response to the word GOOD; subsequently, such statements as "The Young Pioneer helps his comrade" brought maximum salivation, while such statements as "The Fascists destroyed many cities" brought minimum salivation.

In the case of unconditioned physiological responses to the presence or the representation of the attitudinal object, the basis for inference comes directly from the concept of attitude. Just as all definitions of attitude include beliefs, feelings, and overt behavior as indicators of attitude, so do all definitions, explicitly or implicitly, include physiological responses. It is assumed that the magnitude of the physiological reaction is directly and positively related to the extent of arousal or the intensity of feeling; thus, the greater the physiological response, the stronger and/or more extreme the attitude is presumed to be. Here again, however, there are problems in inferring the nature of the attitude being reflected. Most measures of physiological reaction give direct indications only of the extent of arousal; they do not reveal whether the corresponding emotion is pleasurable or unpleasurable. In general, in attempts to assess attitudes toward social groups via measurement of physiological responses, it has been assumed that the range of affect is not from strongly favorable to strongly unfavorable but rather from accepting, or neutral, to strongly unfavorable; thus the inference has been drawn that the greater the physiological response, the more unfavorable the attitude. If Hess' technique of photographing pupillary constriction-dilation can be adapted to the study of attitudes, it would provide a much firmer basis for inferences about the direction of attitude, since the reaction being measured shows a differential response to pleasant and unpleasant stimuli.

In the case of the conditioned physiological responses, the basis for inference is somewhat different, stemming from learning theory. A response that has been conditioned to a given stimulus tends to generalize to stimuli that are similar. Thus, if a response that has been conditioned to the concept "good" appears when the attitudinal object is presented, the inference is that the subject considers the object good—that is, that his attitude toward it is favorable; if the response does not appear when the attitudinal object is presented, the inference is that the subject does not

consider it good—that is, that his attitude toward it is not favorable.

The purpose of the physiological measures may or may not be apparent to the subject. In the Westie and DeFleur (1959) study, for example, subjects presumably realized that the physiological measures were being used as indicators of their reactions to the interracial pictures. In the Rankin and Campbell (1955) experiment, on the other hand, subjects were led to believe that they were taking part in a word-association study and that it was their GSRs to the stimulus words (rather than to the Negro and white experimenters) that were being investigated. Whether or not the purpose is clear to the subject, the fact that the responses measured are not subject to conscious control would seem to eliminate the possibility of modification of responses in order to present a certain picture of the self.

However, physiological responses may be quite sensitive to influences other than those in which the investigator is interested—both to other aspects of the stimulus material and to other environmental influences. It is difficult to control the experimental situation so completely that other factors are ruled out as possible determinants of the response.

Again, questions such as these point to the need for extreme caution in drawing inferences about the attitude of a given individual from a measure of this type. But, again, they point to encouraging possibilities for empirical research and to the opportunity to greatly increase our understanding of attitudes and their relation to various kinds of response, by the use of instruments yielding different types of evidence.

REFERENCES

Adorno, T. W., Frenkel-Brunswik, Else, Levenson, D. J., & Sanford, R. N. *The authoritarian personality.* New York: Harper, 1950.

Allport, G. W. The historical background of modern social psychology. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. 1. *Theory and method.* Cambridge, Mass.: Addison-Wesley, 1954. Pp. 3–56.

Bagby, J. W. A cross-cultural study of perceptual predominance in binocular rivalry. *J. abnorm. soc. Psychol.*, 1957, **54**, 331–334.

Bass, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.

Bray, D. The prediction of behavior from two attitude scales. *J. abnorm. soc. Psychol.*, 1950, **45**, 64–84.

Brown, J. F. A modification of the Rosenzweig Picture-Frustration Test to study hostile interracial attitudes. *J. Psychol.*, 1947, **24**, 247–272.

Campbell, D. T. The indirect assessment of social attitudes. *Psychol. Bull.*, 1950, **47**, 15–38.

Campbell, D. T. Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 6. *Investigations of man as socius: Their place in psychology and the social sciences.* New York: McGraw-Hill, 1961.

Criswell, Joan H. Racial cleavages in Negro-white groups. *Sociometry*, 1937, **1**, 87–89.

Criswell, Joan H. Social structure revealed in a sociometric retest. *Sociometry*, 1939, **2**, 69–75.

Cronbach, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, **6**, 475–494.

Cronbach, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, **10**, 3–31.

DeFleur, M. L., & Westie, F. R. Verbal attitudes and overt acts: An experiment on the salience of attitudes. *Amer. sociol. Rev.*, 1958, **23**, 667–673.

Deri, Susan, Dinnerstein, Dorothy, Harding, J., & Pepitone, A. D. Techniques for the diagnosis and measurement of intergroup attitudes and behavior. *Psychol. Bull.*, 1948, **45**, 248–271.

Edwards, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.

Edwards, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden Press, 1957.

Getzels, J. W. The assessment of personality and prejudice by the method of paired direct and projective questions. Unpublished doctoral dissertation, Harvard University, 1951.

Hammond, K. R. Measuring attitudes by error-choice: An indirect method. *J. abnorm. soc. Psychol.*, 1948, **43**, 38–48.

HESS, E. H., & POLT, J. M. Pupil size as related to interest value of visual stimuli. *Science*, 1960, **132**, 349–350.

HOROWITZ, E. L., & HOROWITZ, RUTH E. Development of social attitudes in children. *Sociometry*, 1938, **1**, 301–338.

HOVLAND, C. I., & SHERIF, M. Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, **47**, 822–832.

JONES, E. E., & KOHLER, RIKA. The effects of plausibility on the learning of controversial statements. *J. abnorm. soc. Psychol.*, 1958, **57**, 315–320.

KINSEY, A. C., POMEROY, W. B., & MARTIN, C. E. *Sexual behavior in the human male.* Philadelphia, Pa.: Saunders, 1948.

LAPIERE, R. T. Attitudes vs. actions. *Soc. Forces*, 1934, **14**, 230–237.

LAZARSFELD, P. F. Latent structure analysis. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 3. *Formulations of the person and the social context.* New York: McGraw-Hill, 1959.

LEVINE, J. M., & MURPHY, G. The learning and forgetting of controversial material. *J. abnorm. soc. Psychol.*, 1943, **38**, 507–517.

MANIS, M. The interpretation of opinion statements as a function of recipient attitude. *J. abnorm. soc. Psychol.*, 1960, **60**, 340–344.

MANN, J. H. The effect of inter-racial contact on sociometric choices and perceptions. *J. soc. Psychol.*, 1959, **50**, 143–152. (a)

MANN, J. H. The relationship between cognitive, affective and behavioral aspects of racial prejudice. *J. soc. Psychol.*, 1959, **49**, 223–228. (b)

MORENO, J. L. *Who shall survive?* (Orig. publ. 1934) (Rev. ed.) Beacon, N. Y.: Beacon House, 1943.

MORGAN, J. J. B. Attitudes of students toward the Japanese. *J. soc. Psychol.*, 1945, **21**, 219–227.

MUSSEN, P. H. The reliability and validity of the Horowitz Faces Test. *J. abnorm. soc. Psychol.*, 1950, **45**, 504–506. (a)

MUSSEN, P. H. Some personality and social factors related to changes in children's attitudes toward Negroes. *J. abnorm. soc. Psychol.*, 1950, **45**, 423–441. (b)

PETTIGREW, T. F., ALLPORT, G. W., & BARNETT, E. O. Binocular resolution and perception of race in South Africa. *Brit. J. Psychol.*, 1958, **49**, 265–278.

PROSHANSKY, H. M. A projective method for the study of attitudes. *J. abnorm. soc. Pychol.*, 1943, **38**, 393–395.

RANKIN, R. E., & CAMPBELL, D. T. Galvanic skin response to Negro and white experimenters. *J. abnorm. soc. Psychol.*, 1955, **51**, 30–33.

RIDDLEBERGER, ALICE B., & MOTZ, ANNABELLE B. Prejudice and perception. *Amer. J. Sociol.*, 1957, **62**, 498–503.

SEELEMAN, VIRGINIA. The influence of attitude upon the remembering of pictorial material. *Arch. Psychol., N. Y.*, 1940–41, **36**(No. 258).

SHERIF, M. A study of some social factors in perception. *Arch. Psychol., N. Y.*, 1935, No. 187.

SOMMER, R. On the Brown adaptation of the Rosenzweig P-F for assessing social attitudes. *J. abnorm. soc. Psychol.*, 1954, **49**, 125–128.

STANTON, H., BACK, K. W., & LITWAK, E. Role-playing in survey research. *Amer. J. Sociol.*, 1956, **62**, 172–176.

STANTON, H. R., & LITWAK, E. Toward the development of a short form test of interpersonal competence. *Amer. sociol. Rev.*, 1955, **20**, 668–674.

TAYLOR, J. B. What do attitude scales measure: The problem of social desirability. *J. abnorm. soc. Psychol.*, 1961, **62**, 386–390.

THISTLETHWAITE, D. Attitude and structure as factors in the distortion of reasoning. *J. abnorm. soc. Psychol.*, 1950, **45**, 442–458.

VOLKOVA, B. D. Some characteristics of conditioned reflex formation to verbal stimuli in children. *Sechenov physiol. J. USSR*, 1953, **39**, 540–548.

WATSON, G. B. The measurement of fair-mindedness. *Teachers Coll. Columbia U. Contr. Educ.*, 1925, No. 176.

WEISS, W. The effects on opinions of a change in scale judgments. *J. abnorm. soc. Psychol.*, 1959, **58**, 329–334.

WESCHLER, I. R. An investigation of attitudes toward labor and management by means of the error-choice method. *J. soc. Psychol.*, 1950, **32**, 51–67.

WESCHLER, I. R., & BERNBERG, R. E. Indirect methods of attitude measurement. *Int. J. Opin. attit. Res.*, 1950, **4**, 209–228.

WESTIE, F. R. Negro-white status differentials and social distance. *Amer. sociol. Rev.*, 1952, **17**, 550–558.

WESTIE, F. R. A technique for the measurement of race attitudes. *Amer. sociol. Rev.*, 1953, **18**, 73–78.

WESTIE, F. R. & DEFLEUR, M. L. Autonomic responses and their relationship to race attitudes. *J. abnorm. soc. Psychol.*, 1959, **58**, 340–347.

(Received July 22, 1963)

Psychological Bulletin
1964, Vol. 62, No. 1, 56–66

IS BINOCULAR CORRESPONDENCE AND DISPARITY STILL A DOMINANT FACTOR IN BINOCULAR DEPTH PERCEPTION? [1]

J. L. ZAJĄC
Ottawa, Canada

Laws of binocular correspondence and disparity are regarded as bases for explanation of most of the phenomena of binocular depth perception in the near field of vision. Abnormal phenomena such as differences in astigmatism of the 2 eyes, differences in acuity, cases of natural aniseikonia, investigated by the author, confirm the validity of the laws of correspondence and disparity and also of the law of intersection of visual directional lines. The inadequacies of other theories of binocular depth perception are pointed out and the author concludes that they cannot supplant the laws of binocular correspondence and disparity for the great majority of cases of binocular perception of depth in the near field of vision.

Comprehension of binocular vision in general and of binocular space perception in particular came very tardily and developed very slowly, and the laws of binocular correspondence and disparity as main factors in binocular depth perception in the near field of vision became known by piecemeal steps. Singleness of vision with two eyes is a very natural thing. People even nowadays are rarely aware of double images or conscious of inequality of what they see with the two eyes. Lack of awareness of these things does not trouble most people because normal vision functions very adequately and purposefully without any knowledge of laws of vision.

Johannes Müller (1826) was the first to formulate the doctrine of identical points on the two retinas which, when stimulated, brought about single vision of objects. But he did not relate these identical points to the perception of depth.

[1] Author deceased at time of publication of this paper.

The author has expressed his thanks to Harry Helson for his many useful suggestions in connection with this paper and for revising the text and correcting the English.

The fact that the two eyes gave different views of the same object was known in antiquity. Euclid in the third century B.C. (1573) was aware of this and Leonardo da Vinci (1802) observed that with both eyes he could see an object that for one eye was covered by another object. Porta (1558) assumed that we see only with one eye at a time so that vision alternates between the two eyes. Aquilonus (1613) admitted that there are two different sensations of one object in the two eyes which should introduce confusion in the whole image, but "commonsense" makes good for this and we see distinctly objects as wholes. In the eighteenth century Harris (1775) tried to explain relief of objects by what we see by both eyes and Porterfield (1759) produced drawings of an object as seen by each eye in binocular vision. But neither of these students of vision connected consciously these "disparities" of the views of separate eyes with depth.

This connection of differences of the images of the same object with depth was discovered in the artificially arranged conditions of stereoscopy. We owe this to Brewster (1847) and espe-

cially to Wheatstone (1828, 1852) who both constructed apparatuses known as stereoscopes. This was the most important discovery concerning binocular vision and binocular depth. The work of Helmholtz (1924), Hering (1861-64), and Panum (1858, 1859), in particular, made several points of stereoscopy clear, although they differed in the interpretation of the facts of binocular depth perception and of stereoscopic depth in particular. Panum established in simple experiments the following important relations: Panum areas, Panum effect, and the law of intersection of visual lines (called by him an empirical law) which defines the place of formation of stereoscopic images.

The concept of horopter (the name was used by Aquilonus in somewhat different meaning) was regarded as the geometrical locus for all corresponding points with a given fixation point.

Retinal transversal disparity principle was also defined in the sense that temporal disparity yielded the perception of "nearer," the nasal disparity the perception of "farther," and lack of disparity or correspondence the perception of "equidistance." This could be demonstrated by a simple experiment carried out by the author with a "mixed" stereoscopic fusion, that is, a fusion of crossed and uncrossed double images (Zając, 1959a). He qualified this experiment as crucial because observed phenomena are in conformity with the principles of retinal correspondence and disparity; it also showed, that in the near field of view, it is not geometrical distance of objects from the observer but transversal disparity which is responsible for depth and relief of stereoscopic images.

With the problem of correspondence and disparity goes the problem of the various kinds of depth we may experience in various conditions of vision. We have to discriminate between three kinds of experiences of depth: perception, estimation, and impression of depth. Perceptions of depth with binocular vision or that resulting from stereoscopic fusion of disparate images in the near field of vision are concerned with relative, even minute differences in distance of objects or parts of objects (relief), which can easily be measured. Certainty is the main characteristic of this class of depth perceptions. They differ in certainty and steadiness from both perception of depth with monocular vision and with binocular vision in farther parts of the visual field. In the latter case depth perception becomes similar to monocular depth perception. Estimates mainly concern judgments about "absolute," objective, or geometrical distance of objects from the observer and quantitative evaluations of differences of perceived distance with monocular or binocular vision. Uncertainty is the characteristic of these experiences but they can be learned and improved. Impressions of depth are experienced when we look at drawings, pictures, films, photographs, etc. Here other factors intervene, such as perspective (both linear and aerial), lights and shadows, height, overlay, etc., which furnish "cues" to depth in two-dimensional media (Zając, 1959b, 1962a). Failure to discriminate these categories leads to a wrong appreciation of the role of various factors in depth perception.

It is surprising, for example, that Gibson (1950) did not mention that depth resulting from stereoscopic fusion of his Figure 49 (p. 105) is of quite a different kind from depth produced by perspective photographs which illustrate different gradients of texture. In his opinion these should produce perceived depth but according to the view advanced here they are examples of impressions of depth characteristic of two-dimensional pictures. Gradients of texture are in Gibson's view the main source of perception of a continuous recession into depth. But they are not

necessary for perceiving recession into depth in the normal field of vision; they are, however, useful for impressions of depth in pictures and photographs, being one kind of manifestation of perspective. In the near field of vision the gradient of texture which, given an impression of recession in pictures, may produce a reverse perception of depth when binocular disparity plays the main role. This has been observed many times by the author. It should also be noted that Gibson's particular set of retinal gradients of texture cannot be a stimulus to perception of a unique longitudinal surface because with monocular vision there is a whole family of such surfaces variously inclined to the horizontal plane corresponding to any one set of retinal gradients. This is the case of so-called "equivalent configurations" stressed by Ames (1955).

Heinrich (1938) writes that already with one eye we see depth. "Vision with both eyes can *improve* three-dimensionality experienced with one eye; but cannot *create* it. . . . A photograph . . . gains in depth when looked at with one eye. Binocular vision renders it flat." Here three different kinds of experienced depth are involved.

Controversy concerning Correspondence and Horopter

All the laws and relations mentioned above were from the beginning subject to controversy. First of all, retinal correspondence and its consequences—single vision and equidistance—were questioned. Wheatstone, Helmholtz, and others assumed that one can see double with corresponding points of the retinas. But they did not prove that we can see double with corresponding points, only that some disparate points or lines might fuse stereoscopically more easily than assumed corresponding points would fuse. The problem, as to which images actually fuse stereoscopically and which resist fusion, is another matter and has nothing to do with the principles of correspondence and disparity. This question will not be discussed here. Linschoten (1956) committed a similar error in discussing his principal figure where depth effect was perceived due to disparity within the limits of Panum areas. Linschoten also criticized the correspondence principle on the assumption that one point in the field of vision can be seen at the same time in two different directions (so-called disjunction). This was not proved convincingly in his arrangement. The author did not observe the phenomenon of disjunction in Linschoten's figures. The concept of horopter has also been denounced as superfluous or irrelevant by several authors (Gibson, 1950; Ogle, 1959). Ogle has pointed out that "The horopter problem is far from being solved." In this author's opinion the relation of single vision, perceptual equidistance (phenomenal data), and horopter (geometrical locus), represents for the time being the definition of corresponding points (physiological correlates) until corresponding points are defined more explicitly in terms of nerve elements or otherwise.

Retinal Disparity Principle Contested

The most controversial problem concerns the role of retinal disparity in binocular depth perception. Especially critical in this regard are some members of the Gestalt school. The criticism was mainly based on the "laws of Gestalt" and it was stated time and again that neither disparity of certain lines or points or parts, but disparity of whole entities or figures or Gestalten is responsible for depth or relief. Wilde (1950), for example, tried to prove the Gestalt contention by means of two figures to be fused stereoscopically, composed of symmetrically disposed dots with a different number of dots in the two lines. When fused, Wilde ar-

gued, they formed lines of dots oblique to the frontal plane. Neither Linschoten (1956) nor I could substantiate these findings for I saw mostly horizontal lines of dots with a difference in depth only at the ends of the lines similar to the Panum effect according to the law of disparity. Only very rarely did I see what Wilde reported. It was quite different when the lateral distance between points in front of one eye differed from that in front of the other eye. The same result was also found with asymmetrical positions of Wilde's figures with respect to the two eyes. In this case an oblique line of dots was seen with an inclination to the frontal plane in conformity with the disparity principle. Also other figures composed of horizontal lines with full or hollow small circles on both sides of the lines presented by Wilde were not suitable to prove his thesis that not disparities of lines and dots but disparities of whole figures produce stereoscopic depth.

On the contrary, identical rectangles, except that one of the horizontal sides was composed of short lines interrupted by gaps differently spaced, when fused, did not show a figure of a regular, complete flat rectangle as would be expected from the law of simplicity advocated by Wilde and other adherents of the Gestalt school. These were seen as a figure with three fused sides lying on a plane surface with one side in the form of short segments in various inclinations to that plane due to the disparities of the fused pieces and gaps between them.

There is another kind of criticism which might be leveled against the theory that disparities of figures as entities rather than disparities of particular contours, lines, points, or other parts of a figure decide the relief of stereoscopically fused images of figures. This theory does not mention the factors which would unequivocally determine what will be perceived nearer and what farther and also what will be seen *equidistant*. Furthermore, if the difference between Gestalten as wholes is the condition for perceiving depth, why is only the difference of shape in the horizontal effective and not also that in the vertical dimension since it is known that vertical disparity does not produce depth?

Ogle's experiments with the so-called "induced effect" led him to assume that difference in size of images of the two eyes in the vertical dimension while equal in the horizontal dimension yields in certain conditions a depth effect similar to that produced by horizontal disparities. Ogle (1950) draws the conclusion:

> that the phenomenon *is psychologic in origin, though depending on specific physiologic stimuli, and is essentially described as a subjective rotational change in a frame of reference for the entire binocular visual field* [p. 197].

This might be considered as an expression of doubt on the part of this renowned expert in the field of binocular depth perception, as to whether only horizontal disparity is responsible for the kind of depth experienced with stereoscopic vision. In the writer's experiments, objects (sheets of paper) equal in horizontal and unequal in the vertical dimension were presented to the two eyes and even with stereoscopic fusion the depth effect described by Ogle did not occur. Only by using a lens that changed magnification of the image in one eye in the vertical dimension could the observation reported by Ogle be made. With glasses of certain thickness, however, change in the horizontal dimension also occurs although it may go unnoticed, if it does not exceed threshold value. Ogle did not mention this in his analysis of the induced effect obtained with afocal meridional lenses at axis 180 degrees used in his experiments.

Law of Intersection of Visual Lines

This law, called by Panum an empirical law, was contested by Hering

(1861-64) who based his criticism among others on faulty observations concerning places and sizes of after-images. The observations by Hering proved to be unwarranted by experiments carried out by the author (Zając, 1960b). Hering and also Linschoten (1956) have rejected the law of intersection of visual lines. Both writers assumed that they also disposed of the projectionist theory of depth perception which, they thought, was connected with the intersection of visual lines. Hering instead invented the concept of the Cyclopean eye, which, however, is unsuited for defining the place where stereoscopic images are formed. The author's investigations (Zając, 1959a) proved that this law can provide an instrument of prediction not only of distances and places at which stereoscopic images are formed but also of their shapes and reliefs. It must be understood that the places and reliefs derived from the law are geometrical and objective and not such as are perceived. Intersection of visual lines demonstrates the place of formation of stereoscopic images only when a fixation point is placed in their neighborhood or when we converge the eyes directly at the point of intersection, a condition not taken into account by Hering.

Dynamism and Disparity

Some investigators of space perception, mostly of the Gestalt school, have conceived disparity not as some retinal correlate of depth perception but as a kind of "stress" (Koffka, 1935) generated in the brain by disparate retinal images. Here also the question arises as to why, in certain cases, stress produces the quality of perceived "nearer" and in others the quality of "farther." These facts are not explained by this kind of "dynamic" theory. Linschoten (1956) argued for a dynamic theory of binocular depth perception based on "attraction" of images belonging to the two eyes in the binocular field of vision. He dismissed from consideration all the happenings in the organ of sight and thereby retinal disparity also. His speculations, however, did not produce a convincing and unequivocal picture of what really happens. His attraction theory and the tendency in the field of vision to have the greatest possible number of binocularly fused images, seem to be factors of "automatic" rather than "dynamic" character. In this author's opinion these explanations also eliminate such psychological factors as attention, tendency to have a better look at some objects or details when looking around, and effects of walking, working, reading, etc., on depth in the binocular visual field.

Convergence versus Disparity

Convergence of the eyes is one of the factors rather than disparity which some authors believed produced perception of depth. Convergence is supposed to determine the distance from the observer at which objects are seen and it is also assumed to contribute to the perception of relative depth of objects in the field of view. There are two factors which are considered as inherent in convergence for depth perception: the angle of convergence, and proprioceptive-kinesthetic sensations accompanying convergence and changes of convergence. But we do not perceive angles of convergence and can only roughly estimate their size. Kinesthetic sensations accompanying changes of convergence are very vague and it is impossible to measure their intensity, something that would be necessary to establish any relation between convergence and the relative distance of objects. Many investigators denied the possibility of perceiving relative depth or relief and of estimating the distance from the observer of objects by virtue of the kinesthetic sensations accompanying accommodation and convergence (Zając, 1960c). Nevertheless these sensations have been regarded as

factors in depth perception by many authors (e.g., Gilinsky, 1951; Ittelson & Ames, 1950; Vernon, 1950) and this opinion is incorporated in many manuals of psychology.

Figure-Ground versus Disparity

There is another factor in perception of space and three-dimensionality—figure-ground—which some students of space believe can displace the correspondence and disparity principles. It does so in pictorial representation of space, where, as mentioned before, impressions of depth rather than perceptions are experienced. Arnheim (1954) refers in this connection to the Gestalt laws of integrity and simplicity. His reasoning is as follows: If a line is drawn on a sheet of paper, it seems not to lie in the plane, but on top of it. Were the line seen as lying within the surface of the sheet of paper or of a canvas of a picture, the surfaces would appear broken.

> There is in a plane surface a tendency to maintain its integrity. A line or patch cannot be erased in this manner. Only one avenue of freedom is open—the third dimension. . . . The total pattern splits within the depth dimension.

A pattern on a surface would look in Arnheim's opinion three-dimensional rather than two-dimensional when this would make for a simpler structure and would preserve the integrity of the background. Many observations can be quoted here against this approach. For example, differences in astigmatism (Zając, 1961) or in acuity of the two eyes create situations in which the background (a simple sheet of paper) on which a line or a circle is drawn, is really broken in perception; that is, its parts appear at different distances, and no Gestalt law of integrity or simplicity can prevent this from happening.

I should like to mention also the following observation made under normal conditions of vision: When we are at sea seated near the starboard and are looking to the sea to the right and to the left, we perceive the starboard side as lower than the portside, and vice versa when we are seated by the portside of the boat. The surface of the sea looks broken, the two parts appearing at two different levels. No *Prägnanz* or constancy law can save it from this appearance (Zając, 1962b).

With color stereoscopy a situation may be created where a figure may be seen behind the background; for example, letters printed in black on blue rectangles are perceived behind these rectangles and as if larger in size than letters of equal size on pale gray background. But white letters on blue background look nearer than black letters on white background. No such observations can be made with one eye. This is an example of color stereoscopy which can produce many pitfalls for painters, because the knowledge of color stereoscopy is not sufficiently advanced and is rarely consciously applied by painters. These and other examples show that in many cases figure-ground cannot, even in pictures, overrule the effects of depth due to disparity.

So-Called Empirical Factors versus Disparity

When stereoscopy is involved together with so-called empirical factors, it is difficult to extricate the role of these different factors in depth perception, and proper experiments are difficult to arrange. One must, however, remember that all these factors and their particular role were evolved from initial binocular vision which is mostly used from birth onwards. Empirical factors or cues were thought to oppose or destroy the effects of binocular disparity or correspondence in certain conditions. One observes cases of their supremacy in depth perception. Ogle (1959) thinks

> that in those surroundings that have been artificially produced to provide a conflict between

> stereoscopic stimuli and empirical factors, the meaningless stimuli may be suppressed by meaningful, that is, by perception from the empirical motives for depth [p. 389].

I shall discuss in the following some of these empirical cues. Some experiments with the pseudoscope seem to have proved (Schriever, 1925) that overlay or superposition was an effective factory in overcoming depth effect resulting from disparity. On the other hand, the author's investigations proved that double images formed on the temporal parts of one eye partly covered by images of the other eye or by stereoscopic images appeared as slanting from behind the latter (Zając, 1956). Another example was a case where a stereoscopic image produced by turning a prismatic glass in front of one eye with an oblique movement away from the observer was seen moving across the surface of the background and going behind the part of the background not covered by the prismatic glass. The screen forming the background was then perceived nearer than the stereoscopic image (Zając, 1962c). This is also an example against generalization of Arnheim's law of figure-ground in cases where binocular vision is involved.

Another factor is height which can influence depth perception in monocular vision. In our experiments on movement (Zając, 1962c) it determined the direction of perceived cyclical movement and corresponding depth phenomena, but had no influence whatsoever on depth in binocular vision and could not compete with the factor of disparity. We are, however, accustomed to see farther things or farther parts of a level field as higher, mounting to a horizon line at the height of our eyes. A receding level road is seen mounting in perception when we are in an erect or semierect position, as we normally are, when walking, driving, or even flying. It is one of the laws of perspective. But at near distances with binocular vision, disparity usually prevails over the height factor. The size factor, too, does not seem to counteract effects of disparity (Ittelson, 1960).

> [Only] continuous size change seems to dominate the binocular disparities to a great extent . . . even under conditions optimally favoring disparity, continuous size change produces considerable apparent movement. This, of course, is the predominant cue producing the appearance of movement in motion pictures and television [p. 160].

To this statement of Ittelson's some remarks should be added. When the size of an image changes, accommodation of the eye alters and consequently disparity also may change. We therefore must be cautious in drawing conclusions as to which of the factors—size or disparity—prevails. It should also be stated that in motion pictures and in television we have to do with what we have called impressions of depth. Most of the screen, being a flat plane surface, represents the area of binocular correspondence with only slight disparity, and correspondence is overcome easier by so-called empirical factors than disparity proper.

With pseudoscopic fusion not all depth due to disparity is normally destroyed; rather mixed and unclear images are perceived with parts of figures fused into incoherent wholes. One cannot prove that here whole figures are fused to produce depth or relief, only in accordance with Gestalt laws, contrary to the disparity principle.

Abnormal Cases in Support of the Correspondence and Disparity Principles

Support for the role of binocular disparity in depth perception is found in abnormal cases which may be crucial for proving or disproving certain hypotheses and theories.

In the author's case different astigmatism in the two eyes, great differences of resolving power of the eyes, and a

cataract developing in his right eye, were the abnormal conditions of vision under which his observations were made. Differences in astigmatism of the two eyes have brought about depth effects similar to those resulting from normal disparity (Zając, 1961). When at certain distances from the eyes one eye sees a vertical line as a blurred surface of some 1.5 millimeters width and the other eye perceives it as a distinct but faint line, with both eyes, a blurred surface is seen in an oblique position with reference to the frontal plane. The source of this phenomenon is the same as that of the so-called Panum effect and it is related to the disparity principle. A similar effect giving rise to relief in perception of a surface on which a line or circle is drawn has already been noted. These cases are convincing because the observations were made directly in the normal field of view without using artificial means or aids.

Unexplained Facts

There are certainly many instances that are not sufficiently or directly explained by the correspondence and disparity principles.

In addition to observation made by several authors (Köhler & Emery, 1947; Lewin & Sakuma, 1925; Ogle, 1950), the author can also cite phenomena that cannot be explained directly by binocular correspondence and disparity. Some were observed with afterimages (Zając, 1960b). Thus there was a difference between depth seen in the primary image and the negative afterimage in a complementary color. The afterimage which was seen as much brighter was also perceived much nearer than the original object.

Another observation provisionally unexplained by correspondence and disparity was made by the author in connection with the Witte-König effect (Helson & Wilkinson, 1958; Zając, 1962b). When two short horizontal lines separated by a gap are placed in front of one eye and only one segment not very different in length from the total length of the other two is presented to the other eye, with stereoscopic fusion, the following is observed: the shorter the gap between the two segments, the more frequent and more steady is the maintenance of the gap perceived in the stereoscopic image, and when some lower limit for the length of the gap is reached, closure never occurs. The assumption is that this is due to contrast. The depth phenomenon observed in connection with the Witte-König effect was as if the piece of the continuous line, which disappeared in perception, still played its role when stereoscopic fusion occurred. This seems to contradict the disparity principle. In further investigations, however, when the vision in the author's right eye greatly deteriorated due to developing cataract, it was found that even though the image in one eye was very faint and hardly noticeable, definite distance and relief were perceived. This seems to indicate that even the parts suppressed by contrast, which were below threshold, still contributed in binocular vision to perception of relief.

There are other phenomena, more general in scope, which so far cannot be explained by the principles of correspondence and disparity. They include changes of relative depth, presumably without changing disparity or correspondence, and they occur mostly during prolonged fixation on one point. With steady fixation there is steady increase of depth in stereoscopic vision until some maximum is reached. Then the changes may take the form of a flattening of the whole field of vision when the images of objects or stereoscopic images seem to be nearing the plane of the fixation point. Also "inversions," changes from "nearer" to "farther," and vice

versa, may occur. To the best of the author's knowledge these phenomena cannot be unequivocally explained by any theory of depth perception. It is, however, to be noted, that changes and various processes connected with duration occur in all domains of sensory perception (color, illumination, size, sound, smell, taste, pain, cold, and warm), all within certain limits, and also in other psychological experiences such as affects, emotions, drives, memories, etc. This might be a phenomenon of a general character which follows laws of the living organism in general and of perception in particular. These changes may be an expression of changes in adaptation level (Helson, 1959).

Role of Accommodation and Convergence

Many of the enigmas and doubts concerning depth wtih binocular vision could be solved if proper assessment or measurement of pupils and accommodation could be made when observations are carried out because changes in accommodation and different states of accommodation in the two eyes produce changes in relative sizes, distinctness and positions of images on the retinas, and thus may contribute to changes in perceived depth. An inverse interpretation of the relation of accommodation and apparent distance was given by Ittelson and Ames (1950). The problem for them was to determine whether a change in apparent distance alone could induce a change in accommodation and convergence. The estimates of apparent distance were made with playing cards of different sizes both with static and continuous size changes. The measurements showed changes in accommodation and the conclusion was drawn that accommodation undergoes changes with altered estimates of apparent distance. Size, however, was not taken into account here; but it is a known fact that objects of different sizes or colors require different accommodation and different size of pupil to be put into focus. I think therefore that the problem posed was a false problem and the conclusions were unwarranted.

It is different with convergence. If we have two figures to be fused stereoscopically, we must converge our eyes to the objective distance at which the stereoscopic image is formed. This, however, cannot be called apparent distance. Similarly, when we want to see an object or a point in the normal field of vision distinctly with both eyes, we converge our eyes on the objective distance of that object.

Heinrich (1923, 1938) assumed that accommodation of the eye(s) is an instrument of attention, which, when directed to an object or a point in space, brings about automatically an adequate state of accommodation by which (within the range of accommodation) a sharp, distinct image is formed on the retina(s).

Convergence plays its great role in binocular depth perception but not through angle of convergence or by proprioceptive sensations. When our eyes are converged on a fixation point, this normally determines accommodation of the two eyes and thus also the size and distinctness of the images. The fixation point also defines the places on the two retinas of images of other objects in the field of vision and their distances from the images of the fixation point. It thus determines whether their position is temporal, nasal, or corresponding; in other words, it defines the character and the amount of disparity and resulting depth perception. The role of convergence is important for perception of relative depth, but not of absolute depth.

Because of the role of accommodation and convergence in depth perception, one can divide the whole field of vision

into three areas: (*a*) an area within the range of accommodation and convergence, in which discrimination of small differences of distance and of relief is good; (*b*) an area within the range of practical convergence and outside the range of accommodation in which bigger differences in distances of objects are still perceived but discrimination of relief is weaker; and (*c*) an area outside the range of accommodation and convergence. In the first two areas there are differences in depth perception of what we see with both eyes and with one eye; in the third area there is no difference between binocular and monocular depth perception. Limits between these areas are different for different eyes and different corrective lenses when they are worn.

Conclusions

Where do we stand with regard to binocular correspondence and disparity as correlates of depth perception in binocular vision?

It seems first of all to be proved by many facts (Senden, 1932, 1960) that we see everything as outside in the field of vision, in front of us, even what happens inside the eyeballs or on the retinas, like *mouscae volitantes,* afterimages, etc. We see in front of us in the dark, when we are asleep (dreams), and when we close our eyes. It is also proved by many facts that perceptions of nearer and farther are related in the near field of binocular vision to temporal or nasal disparities of images on the retinas or to the corresponding processes resulting from projection of retinal processes to the visual cortex of the brain. Lack of disparity, or correspondence, is related to the perception of equidistance. The apparatus of correspondence and disparity of images in the two eyes is an innate and stable organization ready for us from the early days of life.

Disparity works not only in the case of fully fused stereoscopic images within the limits of Panum areas, but also, with double images, within certain limits. In the latter case this depth effect is different as concerns stability and certainty.

Intersection of visual directional lines determines the geometrical position of stereoscopic and binocular images in general when we converge our eyes on the point of intersection or on a fixation point placed in its neighborhood. Directional visual lines are not lines of the same kind as lines along which light penetrates into the eyes, for directional lines do not seem to be refracted; they seem to originate at the place of stimulation on the retina and proceed straight through the point of last refraction to the outside. The law of intersection also involves the law of visual angle and defines geometrical sizes and shapes of stereoscopic images (Zając, 1959). It can also serve as an instrument of prediction in stereoscopic vision.

The general conclusion resulting from the present discussion is that so far not any of the proposed theories or hypotheses can relegate or supplant binocular disparity and correspondence, or the law of intersection of visual lines, as correlates of binocular depth perception in the near field of view and as instruments of prediction in stereoscopic vision.

References

Ames, A., Jr. *An interpretative manual for the demonstrations in the psychology research center.* Princeton: Princeton Univer. Press, 1955.

Aquilonus, F. *Opticorum libri sex.* Antwerp, 1613.

Arnheim, R. *Art and visual perception.* London, England: Faber & Faber, 1954.

Brewster, D. On the knowledge of distance given by binocular vision. *Phil. Mag.*, 1847, 30, 315–318.

Euclid. *La prospetivo de Euclide.* (Trans. by R. P. M. Egnatio Danti) Florence, 1573.

GIBSON, J. J. *The perception of the visual world.* Boston, Mass.: Houghton Mifflin, 1950.

GILINSKY, A. S. Perceived size and distance in visual space. *Psychol. Rev.,* 1951, **58,** 460–482.

HARRIS, J. *A treatise of opticks.* 1775.

HEINRICH, W. Le problème de la méthode: Le problème psychologique de l'espace. [The problem of the method: The psychological problem of space.] *Trav. Lab. Psychol. Exp. U. Cracovie,* 1923, 3–44.

HEINRICH, W. Considerations methodiques generales. [General methodical considerations.] *Trav. Lab. Psychol. Exp. U. Cracovie,* 1938, **2,** 152–157.

HELMHOLTZ, H. v. Helmholtz's treatise on physiological optics. In, *Handbuch der Physiologischen Optik.* Vol. 2. (3rd ed. trans. by J. P. C. Southall) New York: Optical Society of America, 1924. Pp. 311–312.

HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. New York: McGraw-Hill, 1959.

HELSON, H., & WILKINSON, A. E. A study of the Witte-König paradoxical fusion effect. *Amer. J. Psychol.,* 1958, **71,** 516–320.

HERING, E. *Beit. Physiol.,* 1861–64, 1–4.

ITTELSON, W. H. *Visual space perception.* New York: Springler, 1960.

ITTELSON, W. H., & AMES, A., JR. Accommodation, convergence and their relation to apparent distance. *J. Psychol.,* 1950, **50,** 43–62.

KOFFKA, K. *Principles of Gestalt psychology.* London, England: Harcourt, Brace, 1935.

KÖHLER, W., & EMERY, D. A. Figural aftereffects in the third dimension of visual space. *Amer. J. Psychol.,* 1947, **60,** 159–201.

LEONARDO DA VINCI. A treatise on painting. (Trans. by J. E. Rigaud) London, 1802.

LEWIN, K., & SAKUMA, K. Die Sehrichtung monokularer and binokularer Objekte und Zustandekommen des Tiefeneffektes. *Psychol. Forsch.,* 1925, **6,** 203–357.

LINSCHOTEN, J. *Strukturanalyse der binokularen Tiefenwahrnehmung: Eine experimentale Untersuchung.* Groningen: Wolters, 1956.

MÜLLER, J. *Zur vergleichenden Physiologie des gesichtssinnes des Menschen und der Tiere.* Leipzig, 1826.

OGLE, K. N. *Researches in binocular vision.* Philadelphia, Pa.: Saunders, 1950.

OGLE, K. N. Theory of stereoscopic vision. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. New York: McGraw-Hill, 1959.

PANUM, P. L. *Physiologische Untersuchungen über das Sehen mit zwei Augen.* Kiel, 1858.

PANUM, P. L. Die scheinbare Grösse der gesehenen Objekte. *Arch. Ophthalmol., Berlin,* 1859.

PORTA, S. B. *Magiae naturalis.* Rome, 1558.

PORTERFIELD, W. *A treatise on the eye: The manner and phenomena of vision.* 1759.

SCHRIEVER, W. Experimentelle Studien über stereoakopisches Sehen. *Z. Psychol.,* 1925, **96,** 113–170.

SENDEN, M. v. *Raum und Gestaltauffassung bei operieten Blindeboren vor und nach der Operation.* Leipzig, Germany: Barth, 1932.

SENDEN, M. v. Space and sight. (Trans. by Peter Heath) London, England: Methuen, 1960.

VERNON, M. D. *A further study of visual perception.* Cambridge, England, 1950.

WHEATSTONE, C. Contribution to the physiology of vision: Part 1. On some remarkable and hitherto unobserved phenomena of binocular vision. *Phil. Trans. Roy. Soc. Lond.,* 1828, 371–394.

WHEATSTONE, C. Contribution to the physiology of vision. Part 2. *Phil. Mag.,* 1852, 504–523.

WILDE, K. Der Punkreiheneffekt und die Rolle der binokularen Querdisparation beim Tiefensehen. *Psychol. Forsch.,* 1950, **23,** 223–262.

ZAJĄC, J. L. La localisation en profondeur des images doubles. [The localization in depth of double images.] *Trav. Lab. Psychol. Exp. U. Cracovie,* 1923, 45–101.

ZAJĄC, J. L. Depth perception of double images in the vicinity of other images. *Acta psychol., Amsterdam,* 1956, **12,** 111–129.

ZAJĄC, J. L. Depth perception of stereoscopic images resulting from fusion of crossed and uncrossed double images. *Amer. J. Psychol.,* 1959, **72,** 163–183. (a)

ZAJĄC, J. L. Z problematski widzenia przestrzennego. [Concerning the problem of space perception.] *Przegl. psychol.,* 1959, **3,** 127–149. (b)

ZAJĄC, J. L. Convergence, accommodation and visual angle as factors in perception of size and distance. *Amer. J. Psychol.,* 1960, **73,** 142–146. (a)

ZAJĄC, J. L. Spatial localization of after-images. *Amer. J. Psychol.,* 1960, **73,** 505–522. (b)

ZAJĄC, J. L. Depth perception and astigmatism. *Amer. J. Psychol.,* 1961, **74,** 133–134.

ZAJĄC, J. L. Przestrzen w obtrazach. [Space in pictures.] *Estetyka,* 1962, **3,** 103–130. (a)

ZAJĄC, J. L. Stereoscopic depth phenomena with Witte-Konig effect. *J. Psychol.,* 1962, **54,** 153–157. (b)

ZAJĄC, J. L. Some investigations on perception of movement and related depth phenomena. *Brit. J. Psychol.,* 1962, **53,** 117–128. (c)

(Received July 24, 1963)

Psychological Bulletin
1964, Vol. 62, No. 1, 67–69

PROPORTION OF LIGHT TO CYCLE AND CRITICAL FLICKER-FUSION FREQUENCY:

A REPLY TO BARTLEY

ADRIENNE THROSBY
University of Sydney

Bartley's clarification of his model relating proportion of light to cycle (P_L) and critical flicker-fusion frequency (CFF) raised further issues. In a previous report, Bartley predicted 3 transitions from flicker to fusion or fusion to flicker as P_L is increased at constant cycle time. Throsby had shown that there was no consistent evidence to support this expectation and in Bartley's latest statement of the model, only 2 transitions were expected; there was no mention of the previous expectation, nor apparently any alteration to the propositions of the model. The restatement of the model also failed to specify luminance values and cycle times, both important determinants of the number of transitions and both necessary for independent testing of the model.

Bartley (1963) aimed to set out a clarification of his model relating proportion of light to cycle (P_L) and critical flicker-fusion frequency (CFF). This was in reply to part of a review of this determinant by Throsby (1962) in which Bartley's model, as then given, was indicated as being inconsistent and not supported by experimental evidence. In Bartley's clarification, the expectations of the model appear to have changed again, as will be shown below.

Bartley has pointed out that more than one P_L can give the same CFF when other stimulus variables are held constant. This can be shown in two ways. The first procedure obtains the critical curves from studies of CFF at various fixed values of P_L (e.g., Lloyd & Landis, 1960; see Throsby, 1962, for further references). From these curves, the number of transitions expected if P_L is varied at a constant cycle time can be inferred. This procedure was called "inferential" by Bartley and Nelson (1961) and includes many more than the Methods 1 and 2 described by Bartley (1963).

The second procedure varies P_L at constant cycle time and notes transitions from flicker to fusion or fusion to flicker. This was called the "direct" method by Bartley and Nelson (1961) and Method 3 by Bartley (1963). The direct method shows that if fusion is reported for a fixed cycle time and P_L, then increasing P_L may reintroduce flicker, which Bartley (1963) claimed as contrary to "conventional expectation [p. 507]." If this general statement, that there may be more than one transition, by either direct or inferential methods, were all that is claimed for Bartley's model, then there would be no dispute.

However, Throsby (1962) showed that Bartley's predicted number of transitions varied from article to article. Three transitions were first expected (Bartley, 1958; Bartley & Nelson, 1960a, 1960b, 1961). At a fixed luminance and cycle time, very low P_L values were always expected to produce flicker. As P_L was increased, this would change to fusion, then to flicker and finally to fusion, giving three transitions. Bartley and Nelson (1961) used the direct method for a "decisive theo-

retical test of the model [p. 41]," and found only two transitions at three of the four luminance levels used. They then concluded that "it is only at very high intensities that short pulses produce flicker [p. 45]." Yet in the subsequent clarification of the model, Bartley (1963) made no mention at all that short P_L values produce flicker. He then expected only two transitions, fusion to flicker to fusion, while for "a longer cycle time" only one transition, from flicker to fusion, was expected.

Bartley and Nelson (1961) also stated that their failure to confirm the earlier expectation of three transitions might have been "because we were not able to use short enough photic pulses; otherwise it is contrary to our expectations [p. 44]." However, Bartley (1963) did not mention this failure in quoting the findings and stated only that they confirmed his more general expectation of more than one transition. The prediction of three transitions has been omitted from the latest version of the model, presumably without any alteration to the propositions of the model.

Throsby (1962) pointed out that there was no consistent evidence to support the expectation of three transitions. This may not be a "devastating criticism," as it was called by Bartley (1963, p. 507), but lack of empirical support must bear on the first of the dimensions of a model which he set out, namely, "how well it handles the facts it presumes to deal with [p. 507]." Bartley (1963) commented that "Throsby's critique throughout considered only the first dimension [p. 507]." Since this dimension was not satisfied by Bartley's earlier model, as given in his previous publications, discussion of his second and third dimensions would have been fruitless.

Bartley (1963) further confused the basic issues by his comments on Throsby's (1962) Figure 3, which showed schematic curves for his two methods as the model then stood. A horizontal line representing a fixed cycle time cut the CFF curves at three points, A, B, and C. Bartley (1963) stated that Throsby "fails to see that the horizontal line . . . is not a CFF curve; thus her statement about the meaning of Points A, B, and C is irrelevant [p. 507]." This is a misrepresentation; comparison of Bartley's paragraph on this point with Throsby's statements (1962, p. 513) shows no difference in interpretation. The significance of data would be clarified in future reports if the ordinate of graphs were simply labeled "frequency" instead of "CFF." A horizontal line would then clearly be a certain frequency and the threshold curves or transition points, by any method, would show critical values for the given combination of stimulus conditions of area, luminance, etc. Any combination of frequency and P_L below the critical curve would result in reports of flicker, and above the curve in reports of fusion.

One of the main criticisms to be made of Bartley's clarification of his model is the lack of specification of the stimulus conditions under which it is expected to hold. The cycle times required for a given number of transitions appear to be arbitrarily selected and yet small variation from the actual cycle time chosen can alter the number of transitions obtained. For example, if in the top curve of Figure 2 in the paper by Bartley and Nelson (1960b), the horizontal line representing fixed cycle time were lowered only 5 cycles per second, the number of transitions would change from three to one. Cycle time was not specified at all by Bartley (1963) for the main statement of the model with two expected transitions.

One transition is expected only "were a longer cycle time chosen [p. 507]," "longer" being undefined.

A second important but unspecified stimulus condition is luminance, previously named by Bartley and Nelson (1960a) as "one of the factors determining whether there is more than one (transition) [p. 242]." The prediction of three transitions at very high luminances, discussed above, was omitted in the clarification; there was not even a general reference to luminance, although from the shape of the CFF-P_L curves it is directly relevant.

An additional criterion for a model, implicit in Bartley's three dimensions (1963, p. 507), would surely be the precision with which the model specifies the conditions under which its expectations are fulfilled. This clearly has an important bearing on its predictive powers and unless at least some conditions are specified, it is incapable of being tested by anyone other than its author. Bartley (1963) claimed that "the model has inherent in it a number of further predictions which merit testing [p. 507]," but these have not yet been given.

REFERENCES

Bartley, S. H. Some factors influencing critical flicker frequency. *J. Psychol.*, 1958, **46**, 107–115.

Bartley, S. H. Clarification of Bartley's model: A reply to Throsby. *Psychol. Bull.*, 1963, **60**, 505–508.

Bartley, S. H., & Nelson, T. M. The equivalence of various pulse-to-cycle fractions in producing critical flicker frequency. *J. Opt. Soc. Amer.*, 1960, **50**, 241–244. (a)

Bartley, S. H., & Nelson, T. M. Some relations between pulse-to-cycle fraction and critical flicker frequency. *Percept. mot. Skills*, 1960, **10**, 3–8. (b)

Bartley, S. H., & Nelson, T. M. A further study of pulse-to-cycle fraction and critical flicker frequency: A decisive theoretical test. *J. Opt. Soc. Amer.*, 1961, **51**, 41–45.

Lloyd, V. V., & Landis, C. Role of the light-dark ratio as a determinant of the flicker-fusion threshold. *J. Opt. Soc. Amer.*, 1960, **50**, 332–336.

Throsby, Adrienne. Proportion of light to cycle as a determinant of critical flicker-fusion frequency. *Psychol. Bull.*, 1962, **59**, 510–519.

(Received February 3, 1964)

Psychological Bulletin
1964, Vol. 62, No. 1, 70–72

FURTHER CLARIFICATION OF BARTLEY'S MODEL

S. HOWARD BARTLEY
Michigan State University

A reply to Throsby's earlier criticism of the steps taken by Bartley to clarify the relations between CFF, PCF, and the basis for flicker and fusion. By distinguishing between fact, theory, and model it is shown that there has been no violation with regard to the function of each of these in interpreting data, in the application of the model, or in formulation of theory.

I am surprised that my reply to Throsby's original critical article did not make matters clear. Her criticisms of my position arise from a failure to distinguish between the functions of empirical fact, theory, and model. I will try to disentangle these as they concern the problem of proportion of photic pulse to cycle in critical flicker-fusion frequency (CFF).

1. Years ago (Bartley, 1937), I observed that the members of a family of curves (each for a separate pulse-to-cycle-fraction [PCF]) representing plots of CFF against Log I, were not parallel. This was an empirical finding and was more recently corroborated in subsequent experiments.

I inferred from graphs that where any two PCF curves crossed, the point of intersection could be taken as meaning that the two curves (thus the two PCFs) were equivalent under the CFF and Log I conditions in the graph. This led to a second type of inference: once CFF is reached under a given set of conditions, that is, with a given intermittency cycle length and a given Log I, lengthening the photic pulse (raising PCF) till the whole cycle would be filled with the pulse, would not maintain fusion throughout. Flicker would reappear when the second of the two PCF values was reached. This is totally contradictory to traditional expectations and I believed I had hit upon something new—and important. This was the beginning of model construction.

2. In a second step (Bartley & Nelson, 1960), data were obtained to plot CFF against PCF. Others had done the same thing, and three general forms of curve resulted. Some obtained a gradually descending linear curve; others obtained linearly ascending curves. Still others obtained bow-shaped curves which began low, ascended, and later descended. We were among those obtaining the bow curve. This variety puzzled a number of workers. Landis (1954) attempted to account for it, but did not succeed.

Bartley (1958) showed from the facts at hand, that intensity manipulation alone could produce the three forms of curves just mentioned. This assertion, though capable of paper-and-pencil demonstration, was not equivalent to empirical production of all three sorts of curves by actually manipulating the factor of intensity. The demonstration was produced only by transformation from three PCF curves (in which CFF was plotted against Log I) to form a graph in which CFF was plotted against PCF. This was a further application of the model, and as such was fulfilling a perfectly legitimate and useful function.

3. More recently, Bartley and Nelson (1960) made certain inferences from one of these graphs. We drew several horizontal lines across the graph, each from a given CFF. Some inter-

sected the PCF curve at two places, one at only one point, depending upon the CFF used, and thus upon the cycle length chosen. This was an inferential device (model) to show certain expectations, namely, that at least two transition points between flicker and fusion could be expected under certain fixed conditions.

As in my 1958 study, we did extend the model by saying that under conditions producing two transitions, a third one would be expected were the bow curve extended beyond the conditions actually used. This was based on certain reasonable suppositions and on certain neurophysiological findings of Granit and Thurman (1934, 1935). At this point, fact, theory, and model were intertwined.

4. Next, Bartley and Nelson (1961) showed, by actual experiment, that two transition points could be obtained empirically. Our apparatus did not provide for the full extension of conditions to determine empirically that a third transition between flicker and fusion could occur. Let it be understood that to criticize the failure to obtain the third point is not what it appears to Throsby to be. One is allowed to project (a function of a model) and experiment. If perchance, one fails to corroborate the projection (prediction), this does not topple the previous empirical facts. All that negative critics need to do is to show by experiment, not by something less, that the third transition which bothers Thorsby cannot take place. Such a demonstration would not invalidate the two transitions already demonstrated. The fact of two transitions, is alone so foreign to traditional expectations as to constitute a useful discovery.

5. It was the model and the discovery that led to theory. My theory for several years has been that the off-response in the visual system's reaction to photic radiation would account for the second transition, that is, the transition from fusion back to flicker if pulse length (the PCF) of the cycle was increased. I based this on Bartley and Bishop's (1940, 1942) earlier work in recording from the optic nerve. They (we) found that short photic pulses do not produce sizable (if any) off-responses. Off-responses are produced only after pulse durations reach a sizable value. Throsby has produced no evidence, nor has she tried, to form a more plausible substitute to account for the emergence of flicker. Instead, she seems to feel that the assertions I have made regarding a return to flicker as PCF is increased are either not true or are irrelevant.

6. We can come back now to further discussion of the possible third transition between flicker and fusion. This concerns the theory (not the model). Granit and Therman (1934, 1935) showed that a photic pulse producing an off-discharge could have its off-discharge inhibited if a second pulse was delivered soon afterwards. Thus, as PCF lengthens, that is, as the pulse in each cycle gets longer it is followed sooner by the succeeding pulse (in the next cycle). Hence the off-discharge that was taken to be the signal for flicker is now inhibited, and fusion sets in again as the cycle is further filled with the pulse.

I see nothing assailable in the position I have developed and as I described it. Facts stand, models allow for prediction, and theory is intended to account for the facts, although the latter two are often hard to disentangle.

REFERENCES

Bartley, S. H. The neural determination of critical flicker frequency. *J. exp. Psychol.*, 1937, 21, 678–686.

BARTLEY, S. H. Some factors influencing critical flicker frequency. *J. Psychol.*, 1958, **46**, 107–115.

BARTLEY, S. H., & BISHOP, G. H. Optic nerve response to retinal stimulation in the rabbit. *Proc. Soc. Exp. Biol. Med.*, 1940, **44**, 39–41.

BARTLEY, S. H., & BISHOP, G. H. Some features of the optic-nerve discharge in the rabbit and cat. *J. cell. comp. Physiol.*, 1942, **19**, 79–93.

BARTLEY, S. H., & NELSON, T. M. Some relations between pulse-to-cycle fraction and critical flicker frequency. *Percept. mot. Skills*, 1960, **10**, 3–8.

BARTLEY, S. H., & NELSON, T. M. A further study of pulse-to-cycle fraction and critical flicker frequency. *J. Opt. Soc. Amer.*, 1961, **51**, 41–45.

GRANIT, R., & THERMAN, P. O. Inhibition of the off-effect in the optic nerve and its relation to the equivalent phase of the retinal response. *J. Physiol.*, 1934, **81**, 47–48.

GRANIT, R., & THERMAN, P. O. Excitation and inhibition in the retina and in the optic nerve. *J. Physiol.*, 1935, **83**, 359–381.

LANDIS, C. Determinants of the critical flicker-fusion threshold. *Physiol. Rev.*, 1954, **34**, 259–286.

(Received February 10, 1964)

VOL. 62, No. 2 AUGUST 1964

Psychological Bulletin

BEHAVIOR THERAPY:

A REVIEW [1]

JOHN M. GROSSBERG

San Diego State College

The questionable effectiveness of traditional psychodynamic psychotherapies and the development of brief new treatment techniques derived from modern learning theory have stimulated interest in applications of conditioning procedures to behavior disorders. A review of this literature revealed that behavior therapies have been applied to many neurotic and psychotic disorders, and have been most successful with disorders involving specific maladaptive behaviors. Conditioning procedures were highly effective with phobic reactions, anxiety reactions, enuresis, stuttering, and tics, but disappointing with alcoholism and some sexual disorders. Cures seemed long-lasting, with remarkably little evidence of the symptom substitution predicted by psychodynamic depth theories. Behavior therapy offers promising opportunities for the application of well-established psychological principles to the treatment of maladaptive behavior.

Serious efforts by psychologists to apply the rigorous standards of science to the study of psychotherapy have increased since World War II. The use of clinical and other psychologists in clinical capacities during the war due to the pressing need for therapeutic assistance may have instigated this interest.

Earlier attempts to integrate Watsonian behaviorism and the Pavlovian position with personality variables were essentially translations of the facts of personality behavior into the new language of conditioning. This "dictionary construction" (Eysenck, 1957) has moved ahead rapidly in the last 15 years, although the Russian position that neurosis was caused by actual cerebral pathology (Pavlov, 1941) did not retain many influential adherents among English-speaking psychologists. However, many accepted and retained the concept of a disturbed central state which was manifested in observable neurotic behavior. This notion is also basic to psychodynamic personality theories, as exemplified by Dollard and Miller's (1950) comprehensive translation of psychoanalysis in terms of Hull's behavior theory, where internal drives in conflict are said to produce a variety of maladaptive behaviors called symptoms. Translations such as this have helped to reduce vague psychodynamic terms to behavioral language, furnishing a common ground for prac-

[1] This paper is a portion of Technical Report No. 1, Grant RD-892P-63 from the Vocational Rehabilitation Administration, United States Department of Health, Education, and Welfare, to the Western Behavioral Sciences Institute, La Jolla, California.

The author wishes to thank E. G. Aiken for his helpful comments during the preparation of this paper.

titioners and experimenters. However, they have also served to perpetuate psychodynamic therapies unchanged and have failed to provide new or more effective therapeutic procedures.

Recent investigators (Kanfer, 1961; Krasner, 1958) have regarded psychotherapy as a verbal interaction process in which verbal interbehavior is subject to lawful modification by appropriate reinforcing operations, extinction procedures, and the like, in no way qualitatively different from other behavior. However, while generally rejecting a dualistic central conflict-peripheral symptom model, some dubious relics of psychodynamic thinking remain. These include the assumption that changes in verbal behavior are necessary prerequisites for improvements in other maladaptive behaviors (Krasner, 1955; Metzner, 1961), or that it is important to talk about things you do not want to, make more affect statements, or recall early memories (Greenspoon, 1962). Although it is well established that there is systematic modification of verbal behavior in psychotherapy (Bandura, 1961; Frank, 1961) and experimental verbal conditioning analogues of psychotherapy (Greenspoon, 1962), both Greenspoon and Kanfer (1961) have concluded that there is very little evidence for the transfer of interview effects to other behaviors. Zax and Klein (1960) reviewed the literature on behavior changes following psychotherapy and concluded that there was no good evidence relating verbal behavior changes during therapy to behavioral changes in the family and community. Bandura (1961) indicated that current research on verbal conditioning without awareness casts serious doubt on the value of therapies whose primary aim was the development of insight, and Hobbs (1962) concluded that insight is either independent of personality change or an occasional byproduct of such change.

Although investigators in this area have focused on the direct effects of therapist verbalizations on patient behavior rather than on some central process, we have seen that elaborate verbal exchange is often assumed to be an important part of treating maladaptive behavior, without strong supporting evidence. In the last 15 years, Eysenck, Wolpe, and their students have departed radically from psychodynamic and verbal insight assumptions and their related procedures for modifying maladaptive behavior. Eysenck (1952, 1961) reviewed studies of the recovery rate from neurotic disorders claimed for psychoanalysis and other dynamic psychotherapies as compared with custodial treatment and/or nonspecific treatment. He concluded that psychodynamic therapies were no more successful than the comparison control procedures, and questioned the effectiveness of traditional therapy in producing beneficial behavior changes, stating in effect that he failed to see the Emperor's new clothes. Levitt (1957) reached a similar conclusion after evaluating outpatient child psychotherapy studies.

These findings led Eysenck and Wolpe to reject traditional psychodynamic approaches to treatment. Instead Eysenck (1960b) defined neurotic behavior as maladaptive habits formed through a process of conditioning and capable of being extinguished through several techniques of demonstrated effectiveness in the laboratory. There is no complex, no illness, and treatment is directed entirely to the symptoms, as distinguished from psychotherapy with its stress on hypothetical underlying complexes and disease processes. Wolpe (1958) proposed a similar view, that neurosis is persistent maladaptive learned behavior with a prominent

anxiety component, acquired in anxiety-generating circumstances.

Skinner (1953) also attacked the assumption that behavior disorders were a reflection of some internal conflict or sickness. He termed the conversion of disordered behavior into a thing labeled "neurosis" an explanatory fiction, which has had several unfortunate consequences. It has encouraged the therapist to avoid specifying the behavior to be corrected, and by suggesting a single cause for a variety of maladaptive behaviors it has implied a uniformity not to be found in the data. More serious from a humanitarian viewpoint, it has fostered the belief that inner causes must be extirpated or purged, leading to electrical, surgical, or pharmacological assaults of questionable benefit to the person (Gordon, 1948; Reider, 1955). In Skinner's view, motivation does not reside in the person, so that conceptions such as insight, the lifting of repressions, making the unconscious conscious, etc., are unnecessary for behavior change. The idea that people learn to do things because they "want to" is discarded, and the problem becomes one of altering the reinforcing environment so that adequate behaviors are maintained, effective new behaviors learned, and inadequate behaviors extinguished (Michael & Meyerson, 1962).

While Eysenck, Wolpe, and Skinner are in accord with respect to the emphasis on direct behavior modification as the focus of remedial efforts, Eysenck and Wolpe employ a Hullian behavior theory with constructs such as inhibition and drive which are rejected by Skinner. Although this theoretical divergence is of great importance, it is beyond the scope of this paper.

Behavior therapy is derived from the rejection of traditional psychodynamic personality theories, and consists of the application of the principles of modern learning theory to the treatment of behavior disorders. Theorists who have described behavior therapies (Bandura, 1961; Dunlap, 1932; Eysenck, 1960a; Metzner, 1961; Skinner, 1953; Wolpe, 1958) have often used different terms for similar procedures. Descriptive terms for treatment procedures tend to be confounded with the learning mechanisms involved in behavior modification. In this paper, the following classification is used:

1. Aversion therapy. (*a*) Aversive stimuli are presented in conjunction with the elicitation of a maladaptive response (counterconditioning or punishment, Bandura; aversion therapy, Eysenck; conditioning avoidance response, Wolpe). (*b*) Aversive stimuli are reduced or terminated upon the occurrence of a response whose low frequency constitutes a behavior deficit (negative reinforcement, Skinner; conditioned anxiety-relief responses, Wolpe).

2. Negative practice. The maladaptive response is elicited repeatedly and frequently (extinction, Bandura; negative practice, Dunlap; conditioned inhibition, Eysenck; reactive inhibition, Metzner).

3. Operant conditioning. Reinforcing stimuli are presented which increase the frequency of responses whose low frequency constitutes a behavior deficit (reward, Bandura; positive conditioning, Eysenck; operant conditioning, Skinner).

4. Reinforcement withdrawal. Positive reinforcing stimuli are withdrawn following the occurrence of a maladaptive response (extinction, Bandura).

5. Desensitization. Stimuli evoking maladaptive responses are repeatedly presented at low intensities which do not provoke a full-blown response. Stimulus intensity is systematically increased until even high intensities no

longer evoke maladaptive responses (counterconditioning, Bandura; reciprocal inhibition, Eysenck; graded stimulus situations, Metzner; reciprocal inhibition with systematic desensitization, Wolpe).

In this review, maladaptive behaviors are divided into neurotic and psychotic disorders for purposes of convenient communication. Within these broad categories, the cases are organized around the dominant maladaptive behavior treated by behavior therapy. Wherever possible, such clusters of cases are grouped according to standard psychiatric nomenclature. The terms *fear* and *anxiety* are used interchangeably.

Neurotic Disorders

Phobic Reaction

The most widely used and successful behavior therapy for phobias was developed by Wolpe (1958). It is a desensitization procedure based on the assumption that relaxation responses are antagonistic to fear. Therapist and patient construct a ranked list of fear-evoking stimulus situations. The patient is then trained to relax, often while hypnotized. He is then told to visualize the weakest fear situation on the list. After a few trials it can usually be visualized without fear, whereupon the next situation in the hierarchy is visualized. Eventually relaxation responses come to be attached not only to all visualized situations on the list but also to their real-life counterparts. The induced relaxation is said to reciprocally inhibit the fear, which extinguishes upon repetitions of the procedure. Wolpe occasionally used an anxiety-relief procedure in which he administered an electric shock to the forearm which the patient terminated by saying "calm" when it reached unbearable intensity. Patients reported that the repetition of calm, which had been associated with fear/pain cessation, was often helpful in reducing fears. Wolpe reported that 90% of 210 patients treated by such methods were cured or much improved. Of the 210 cases, 135 were termed anxiety reactions, including phobias, so the exact number of cured phobics is unknown.

Desensitization methods similar to this have been very successful in treating fear of small animals (Freeman & Kendrick, 1960; Jones, 1924a, 1924b; Lang & Lazovik, 1963; Lazarus, 1959; Lazarus & Abramovitz, 1962; Lazarus & Rachman, 1957; Weinberg & Zaslove, 1963), fear of the dark (Holmes, 1936), fear of height (Holmes, 1936; Lazarus, 1961), fear of cars (Lazarus, 1959; Wolpe, 1962a, 1962b), agoraphobia (Meyer, 1957; Terhune, 1949; Wolpe, 1958), fear of hospitals and injections (Lazarus & Rachman, 1957; Rachman, 1959), and fear of public speaking (Grossberg, 1963). Meyer and Gelder (1963), however, were unsuccessful with agoraphobics and reported one of the few instances of symptom substitution in the behavior therapy literature. Malleson (1959) reported that instructions to practice an examination fear led to its rapid extinction.

Anxiety Reaction

Many patients Wolpe (1958) cured with desensitization complained of pervasive rather than specific anxiety. Wolpe (1961) described the successful extension of desensitization therapy to problems such as feelings of guilt and devaluation. Other cures of anxiety reactions by desensitization and relaxation were reported by Haugen, Dixon, and Dickel (1958) and Lazarus and Rachman (1957).

Gross Stress Reaction (Combat)

During World War II desensitization was successfully used to reduce exaggerated startle responses to noise, with the presentation of low intensity sounds which were gradually made louder until maladaptive responses extinguished (McLaughlin & Millar, 1941; Rudolf, 1961; Saul, Rome, & Leuser, 1946; Schwartz, 1945).

Sexual Disorders

Rachman (1961) reviewed the recent literature on behavior therapies for sexual disorders and his data are included in this section. Freund (1960) reported that aversion therapy with the nauseant drugs, apomorphine and emetine, was no more successful than conventional psychotherapy with 67 male homosexuals. Oswald (1962) failed to help another homosexual with these drugs, although Max (1935) reported a cure with electric shock as the aversive stimulus. Clark (1963), Oswald (1962), and Raymond (1956) used apomorphine successfully to cure four of five fetishists. Glynn and Harper (1961) and Lavin, Thorpe, Barker, Blakemore, and Conway (1961) used apomorphine aversion conditioning to cure male transvestites although Oswald (1962) reported a failure with this procedure. Aversion procedures apparently yield equivocal results, and the picture is further complicated by the fact that apomorphine alone, without aversion conditioning trials, is used to treat neuroses (Kalinowsky & Hoch, 1961).

Dunlap (1928, 1932) reported that negative practice procedures had cured homosexuality but furnished no details. Stevenson and Wolpe (1960) cured two homosexual men by verbally reinforcing assertiveness which aided them in becoming more aggressive with women. Wolpe's systematic desensitization seems to be more successful than aversion methods, with many cures reported for impotence and frigidity (Lazarus, 1963; Lazarus & Rachman, 1957; Wolpe, 1960), and voyeurism and exhibitionism (Wolpe, 1958), although Bond and Hutchison (1960) were only partly successful in treating a chronic exhibitionist.

Conversion Reactions

Sensorimotor Disorders. Hilgard and Marquis (1940) and Sears and Cohen (1933) used classical conditioning methods to reinstate feeling and movement in two women with functional disorders of the arm and hand. Cohen, Hilgard, and Wendt (1933) and Malmo, Davis, and Barza (1952) reinstated seeing and hearing, respectively, by similar procedures. Brady and Lind (1961) cured a functionally blind man by operant conditioning in which his use of visual cues was systematically reinforced. However, the role of conditioning seems confounded with suggestion effects in these cases. For example, in the Malmo, Davis, and Barza case the patient was emphatically told that the conditioning procedure would restore her hearing, and there were distinct opportunities for such effects in the other cases.

Anorexia Nervosa. White (1959) cured a 5-year-old girl who stopped eating following the death of her father who used to help feed her. Successive approximations to independent eating were instituted during play therapy, beginning with sham eating at doll tea parties, followed by the eating of doll portions of food. Bachrach, Erwin, and Mohr (in press) reinstated eating in a woman whose weight had fallen from 118 to 47 pounds by making reinforcements such as visits, books, music, and TV viewing contingent upon her eating.

Obsessive Compulsive Reaction

Dunlap (1930) and Walton (1960c) used negative practice to cure obsessive thoughts of illness and compulsive handwashing. Bevan (1960) used desensitization to eliminate obsessive thoughts of world destruction in a housewife. Lazarus (1958) cured a man with a handicapping compulsion to recheck his work by suggesting under hypnosis that rechecking aroused great fear which was immediately reduced by leaving the task to go on to another. Wolpe (1958) reported two effective "thought stopping" procedures based on the therapist's shouting, "Stop!" whenever the patient signaled that he was experiencing an obsessive thought.

Enuresis

Jones (1960a) reviewed the physiology, psychology, and treatment of nocturnal enuresis, with emphasis on the bell conditioning treatment devised by Mowrer (1938). Jones cited 15 studies in which this treatment was applied to 1,446 enuretics, with a total of 76% cures and 14% failures. Data for "markedly improved" cases are not included since some authors classified these as cures while others retained this as a separate category. There were no reports of symptom substitution, although most investigators reported beneficial personality changes. This suggests that the other adjustment difficulties often mentioned in conjunction with enuresis are probably secondary to the enuresis. The conditioning procedure appears to be very successful, especially in view of the fact that many of the cases were hard core enuretics who had failed to respond to other treatments. There is some evidence that the cure rate could be further improved by using partial reinforcement, more widely spaced trials, and overlearning (Eysenck, 1963).

Jones (1956) cured a woman with excessive urinary frequency by allowing her to observe her own bladder pressure reading which preceded the urge to urinate. He then introduced increasing amounts of liquid directly into her bladder while manipulating the pressure gauge so that it always read below her urinary threshold.

Alcoholism

To date, some form of aversive procedure has been the sole behavior therapy for alcoholism. Franks (1958), in a comprehensive review of the literature, criticized those who have used aversive methods for their ignorance of learning theory, especially the facts regarding the necessary relationships between conditioned and unconditioned stimuli and responses. In addition, he cited the lack of uniformity of procedures, the dearth of technical details, and general failure to use noxious stimuli other than nauseant drugs, such as electric shock, which would afford more stimulus control and has fewer side effects. The large sample studies cited by Franks reported average total abstinence rates of approximately 50% several years after treatment, which is no better than the recovery rate reported for religious conversion, social therapy, or standard psychotherapy (Carlson, 1944). However, aversion therapy seems to be shorter and less expensive than many of these other treatments.

Speech Disorders

The unsystematic use of gradual approximations to adequate speaking is very common in many areas of speech retraining. With stuttering, most behavioral approaches have used some form of negative practice. Lehner (1954), in a review of negative practice methods, concluded that results

were best in habit residual cases where the pressures that originally produced stuttering are absent, and where the patient wants speech improvement without more general behavior modification. Jones (1955) demonstrated that decreases in stuttering with repetition was a form of experimental extinction, opening the way for learning interpretations of this behavior. Sheehan (1951) and Wischner (1950) proposed similar learning theories of stuttering based on a response conflict model, and Goldiamond (1962) presented an operant conditioning theory. Shane (1955) found that a loud masking noise decreased ongoing stuttering. Cherry and Sayers (1956) and Cherry, Sayers, and Marland (1955) confirmed this and reported that "speech shadowing" practice in which the patient parroted material read or spoken by the therapist led to clinical improvement. Walton and Black (1958) cured a man who stuttered severely during telephone calls, using the shadowing method. Lazarus and Rachman (1957) reported a cure with systematic desensitization, and Flanagan, Goldiamond, and Azrin (1958) decreased stuttering frequency by having each stuttered word produce a brief blast of loud noise. Bloodstein (1949) reviewed and classified the many diverse techniques which successfully reduced stuttering. His paper clearly illustrates the dangers of using therapeutic effectiveness to verify theoretical formulations. He reported that stuttering is often reduced or abolished when the communicative aspects of speech are reduced, as in speech shadowing; when the patient talks to an understanding or helpful person; or with therapists who have a sincere and unbounded faith in their own particular method, whatever it is.

Aphonia. Bangs and Freidinger (1949, 1950) cured an aphonic girl and a woman by requiring graded vocal activities that were made more and more similar to voiced social speaking. Walton and Black (1960) cured an aphonic woman by changing the length of vocal practice sessions to reward voice volume increases and punish volume decreases, and by practice in vocal tasks approximating everyday life conversation.

Special Symptom Reactions

Dunlap (1932) first reported that the negative practice of tics led to their abolition, and this method was successfully employed by Ernest and Jones (Jones, 1960b), Walton (1961), and Yates (1958a). Barrett (1962) reduced tic frequency by having a patient listen to recorded music which was automatically interrupted by each tic movement.

Liversedge and Sylvester (1955; Sylvester & Liversedge, 1960) reviewed history and treatment of occupational cramps, primarily writer's cramp, and built several retraining devices which delivered an electric shock whenever writing practice was disrupted by spasms or tremor. With this form of aversion therapy they cured 24 of 39 patients. Beech (1960) cured three of four patients with negative practice of the cramp or systematic desensitization.

Dunlap (1930, 1932) reduced the frequency of nailbiting and thumbsucking in "many" cases with negative practice although complete extinction was rare since motivation declined along with habit frequency. Baer (1962) reduced thumbsucking frequency in three boys by interrupting movie cartoons each time the thumb was sucked.

Kuehner (1956) cured "several" cases of teeth gnashing during sleep by a posthypnotic suggestion that the patient would awaken when the mastication muscles contracted strongly. Wil-

liams (1959) extinguished bedtime screaming tantrums in a 21-month-old boy by leaving the boy's room and shutting the bedroom door. Zimmerman and Zimmerman (1962) used operant conditioning to eliminate maladaptive classroom behavior in two psychiatric inpatient boys. The teacher withheld attention until adaptive behavior was emitted, and then gave generous social reinforcement, resulting in marked academic and social improvements. While these last two reports seem to describe nothing other than what some parents and teachers have done for generations, nevertheless it is important to demonstrate that there are general psychological principles underlying successful behavior modifications.

Slack (1960) used an operant procedure to establish and maintain regular therapy contacts with highly resistant juvenile delinquents. Initially he tolerated irregular attendance and then gradually shaped regular punctual attendance by reinforcing it with money, food, cigarettes, and occasional bonuses. Slack does not regard this method as therapy but rather as a means of establishing sufficient contact so that the patient can learn that help is available if he wants it.

Miscellaneous Disorders

Walton (1960b) cured a woman whose excessive scratching aggravated a severe skin rash by instructing her fiance and family to omit the excessive attention which seemed to reinforce this response. Walton (1960a) relieved a case of bronchial asthma in an inhibited man whose attacks seemed related to a demanding wife, family, and business associates. After desensitization and assertiveness training, social relationships and asthma improved markedly.

Efron (1957) used stimulus substitution therapeutically to eliminate chronic uncinate grand mal convulsions. After Efron discovered that an aromatic odor inhibited the woman's fits, he was able to train her to substitute looking at a bracelet for sniffing the aromatic substance, thus effecting a conditioned inhibition of the fits. Rubinstein (1931) cured two morphine addicts by associating a neutral stimulus with daily morphine injections, and then gradually reducing the concentration of drug in the injections. Kalinowsky and Hoch (1961, p. 214) reported that Soviet psychiatrists had been moderately successful using a similar stimulus substitution procedure to reduce the amount of insulin required to induce coma in treating schizophrenics.

Psychotic Disorders

Adult Psychosis

Although behavior therapy applied to psychotics has resulted in beneficial behavior changes such as increasing sociability, decreased delusional talk, etc., few investigators have reported hospital discharges attributable to treatment. Operant conditioning procedures were used in the majority of studies.

Fuller (1949) operantly conditioned and extinguished an arm-raising response in a bedfast vegetative idiot, using a sugar-milk solution as reinforcement. Peters and Jenkins (1954) used fudge to reinforce the solving of increasingly complex problems, in an attempt to alter schizophrenic rigidity, and found that their experimental group was granted significantly more hospital privileges. Lindsley (1956, 1960) devised an apparatus and procedure for the experimental analysis and control of psychotic behavior. The patient was seated before a sort of vending machine containing a plunger which, when manipulated, could be made to deliver various reinforcing

stimuli. In this way Lindsley studied the behavioral effects of different reinforcing stimuli, reinforcement schedules, drugs, hallucinations, and psychotherapy. King, Armitage, and Tilton (1960) used a similar apparatus with withdrawn schizophrenics in an attempt to elicit more social behavior and environmental interest, and noted significant improvements in ward behavior ratings.

Ayllon and Michael (1959) successfully used operant conditioning and extinction procedures to eliminate specific disruptive behaviors such as excessive floor scrubbing, newspaper hoarding, and pestering visits to the nurses' station. Ayllon and Haughton (1962) eliminated eating problems requiring excessive staff time and effort in 30 of 32 patients by denying access to the dining hall to patients who did not enter during a brief time period. Ayllon (1963) used similar procedures to eliminate food stealing, towel hoarding, and wearing excessive clothing in another psychotic patient. Isaacs, Thomas, and Goldiamond (1960) reinstated talking in two mute catatonics by reinforcing successive approximations to speech with chewing gum. Rickard, Dignan, and Horner (1960) increased the amount of rational speech in a verbose psychotic man by systematically manipulating the therapist's interest responses to his speech content. This man's improvement was still evident 2 years later (Rickard & Dinoff, 1962). Wickland (1963) assigned his patients tasks which more and more closely approximated independent living and working, and found that 40 of 41 chronic schizophrenic women in this program were released from the hospital and were supporting themselves adequately.

Cowden and Ford (1962) found systematic desensitization successful in helping one of two phobic paranoid schizophrenics. Walton (1960d) cured a schizophrenic man with exaggerated fears of women and social relationships by practice in tasks requiring increasing social contacts with women and others, although the patient's history of relapses suggested that his cure might be short-lived. Thompson and Bielinski (1953) reported that six alcoholics whose psychosis preceded their alcoholism were cured of both disorders after drug aversion therapy.

Child Psychosis

Ferster and DeMyer (1961, 1962) described an operant conditioning apparatus and procedure similar to Lindsley's, consisting of many coin-operated vending machines, which they used to analyze and control autistic children's behavior. They reported that these children did the same things as normal children but much less frequently. Various reinforcing operations were successful in widening the children's narrow behavior repertoires. Ferster (1961) presented a detailed analysis of autistic behavior and those factors relevant to its acquisition and maintenance. Wolf, Mees, and Risley (1963) used operant conditioning methods with a 3-year-old boy to eliminate injurious tantrums and bedtime disturbances, and to train him to wear eyeglasses which were needed to prevent permanent eye damage.

Discussion

Behavior therapies have been most successful when applied to neurotic disorders with specific behavioral manifestations. Desensitization procedures have been successfully applied to the greatest variety of neurotic problems, while therapeutic stimulus substitution has rarely been used. Aversion therapy

has produced disappointing results in alcoholism and sexual disorders, although more successful in treating occupational cramps. Negative practice procedures have proven most helpful in treating stuttering and tics. Operant conditioning procedures have been used most frequently with psychotics but as yet there are few reports of clinical recoveries. On the basis of current evidence, behavior therapy appears to be a superior treatment for phobic reactions and enuresis, and definitely worthwhile with selected individual cases of other disorders.

Many psychologists who are interested in behavior therapy have abandoned traditional depth psychotherapies because of their questionable effectiveness (Astin, 1961; Eysenck, 1952, 1961; Levitt, 1957), their basic assumption of a scientifically unacceptable mind-body dualism (Kantor, 1953; Rotter, 1954; Skinner, 1953), and their tenacious devotion to a poorly defined explanatory system which is made to encompass all past, present, and future human experience. Many advocates of the psychodynamic viewpoint, on the other hand, regard behavior therapy as the result of a naive pragmatism which overlooks or ignores the basic therapeutic agent in its procedures, suggestion, and transference (Rosenzweig, 1954). Successful behavior therapy is criticized as palliative and superficial at best, and possibly harmful through symptom substitution (Bookbinder, 1962). There are also objections to the direct, deliberate, and systematic manipulation of the patient's behavior as authoritarian and antihumanistic (for a discussion of this issue, see Bandura, 1961). These criticisms are discussed below.

Suggestion-Transference. It is not logically possible to prove the absence of an effect such as suggestion in eliminating maladaptive behavior, or that substitute symptoms do not occur. However, assuming that suggestion-transference effects are behavior modifications resulting from some aspect of the therapist-patient interaction and patient expectations, rather than the specific treatment procedure, we can then discuss behavior therapy procedures to see what opportunities they afford for such effects. Carlson (1944) discussed the many opportunities for suggestion effects in the ritualistic aversion therapy séances for alcoholism, and his criticism is supported by the finding that outcome rates are approximately the same for widely different treatment methods. Bloodstein (1949) indicated that stuttering frequency may be reduced effectively by numerous novel procedures, complicating the role of behavior therapy in this disorder. Classical conditioning methods applied to conversion reactions also offered many opportunities for suggestion. However, the suggestion-transference hypothesis receives little support in the remaining studies, particularly the desensitization literature. Behavior therapy has often been a last resort, and since patients have previously received psychotherapy or other treatment, there would have been ample prior opportunity for flight into health. Some of the successes reported for various therapies probably represent such suggestion cures. Behavior therapies are generally brief and straightforward, with fewer departures from conventional social interaction than traditional psychotherapies, offering a less awesome atmosphere for suggestion cures. Modifications in maladaptive behavior during behavior therapy are generally gradual and regular rather than sudden and dramatic, and resemble the response curves obtained during acquisition or extinction of conditioned responses. Wolpe

(1962b), in treating a phobic patient, attempted to omit all activities which might give grounds for a suggestion interpretation. He substituted a medical student therapist for himself for five sessions with no disruption of desensitization, and cited the fact that decreases in fear only occurred during desensitization sessions but not in the long intervals between sessions. Meyer and Gelder (1963) and Wolpe (1958, 1961) make the important point that desensitization only alleviates those phobias that are being treated, but other coexisting phobias remain at high strength, indicating a specific treatment effect. Controlled studies by Lang and Lazovik (1963) and Lazarus (1961) contribute additional evidence that desensitization rather than hypnosis or relaxation alone led to improvements. Thus, general consideration of behavior therapy procedures as well as specific investigations of the role of suggestion in behavior therapy support the conclusion that elimination of maladaptive behavior is probably related to the specific method of treatment.

Symptom Substitution. According to psychoanalytic formulations, maladaptive behavior is a derivative of the conflict between the contradictory impulses striving for discharge and defensive forces. This makes it impossible to discharge tension in the usual way, so tension accumulates, and the defensive system is overpowered by excitation, producing distorted discharges which are symptoms. It follows that treatment which leaves the central conflict unresolved will be unsuccessful, and substitute symptoms are expected. This issue is important not only because of the empirical question of the effectiveness of symptomatic treatment and patient welfare, but because it tests a hypothesis directly derived from psychoanalytic theory, thus bearing on the utility of this aspect of the theory (Eysenck, 1960a; Yates, 1958b). The overwhelming evidence of the present review is that therapy directed at elimination of maladaptive behavior ("symptoms") is successful and long-lasting. Substitute symptoms were reported in two cases of all those surveyed (Lazarus & Rachman, 1957; Meyer & Gelder, 1963), although a number of investigators with psychoanalytic backgrounds were particularly sensitive to this problem. Unfortunately, psychotherapists seem to have stressed the hypothetical dangers of only curing the symptoms, while ignoring the very real dangers of the harm that is done by not curing them.

Bookbinder (1962), in a discussion of this issue from a psychodynamic viewpoint, stated that depth psychotherapy treats neurotic predispositions and reorganizes the character structure, rather than manipulating individual symptoms, and is therefore of wider and more permanent benefit. He also criticized behavior therapists for an overly narrow definition of symptoms, and suggested that subtle substitute symptoms such as ways of relating to others, undesirable mood states, and pathological ways of handling one's internal life may be overlooked. He cited several studies in which severe pathological reactions were precipitated by experimental hypnotic symptom removal or medical treatment of ulcers, ulcerative colitis, and asthma. Bookbinder suggested that since most behavior therapists have dealt with neurotic disorders which are painful or inconvenient, perhaps substitution is more likely in severe conditions which pose a danger to life. This is an interesting empirical question, but present evidence indicates that behavior therapy is seldom followed by additional behavior problems or relapse, although

symptom substitution may be a danger in inadequate or improper psychodynamic psychotherapy or the medical treatment of psychosomatic disorders.

Manipulation of the Patient. Behavior therapy has been criticized for mechanical manipulation of the patient, but there is abundant research evidence of subtle, unintentional therapist manipulation, especially of verbal behavior in dynamic psychotherapies (Bandura, 1961; Frank, 1961; Greenspoon, 1962). If there is agreement that all psychological treatment involves a lawful interpersonal influence process, then objections must be to the kind of manipulation or influence involved in behavior therapy, rather than to manipulation per se.

It is unlikely that substantial numbers of traditional psychotherapists will begin to apply behavior therapies, regardless of their demonstrated economy or success. Many mature psychotherapists are personally and professionally committed to personality theories and treatment methods diametrically opposed to those of behavior therapy. According to Frank (1961), evocative therapies which stress self-knowledge and insight appeal to well-educated people who place a high value on self-knowledge and verbal skill, while directive methods have low status for this group. Hobbs (1962) pointed out that insight and understanding appeal to us as central mechanisms in psychotherapy because of our culture's strong commitment to rationality in problem solving. Behavior therapists do not subscribe to a central conflict, symptom derivative model, and therefore insight or the patient's knowledge of the origin of his problems are extraneous to treatment. It is on the proving grounds of the training institutions and laboratories that behavior therapy will compete with traditional theories and practices for the interest of critical young psychologists. Its initial successes will probably be tempered by later reports, as is so often the case with any treatment innovation. There are too few controlled studies in the behavior therapy literature, and there is an obvious need for comparing directly the effectiveness of traditional therapies and behavior therapy with various kinds of maladaptive behaviors. Behavior therapists too often seem prematurely committed to outmoded or oversimplified versions of learning theory, with little clear connection between the theory and practice (Rotter, 1960). Meanwhile, this continues to be an extremely active research area. A new journal has been launched to report work in this field, Wolpe (1961) has been working at the extension of behavior therapy to additional less circumscribed behavior disorders, and others [2] (Lang & Lazovik, 1963) are investigating confounded variables in the desensitization procedure. The demonstrated success of behavior therapy has created challenging new possibilities for the highly effective treatment of behavior disorders and the fruitful marriage of basic science and practice in psychology.

[2] Beatrice Ashem, personal communication, 1963.

REFERENCES

Astin, A. The functional autonomy of psychotherapy. *Amer. Psychologist,* 1961, **16,** 75–79.

Ayllon, T. Intensive treatment of psychotic behavior by stimulus satiation and food reinforcement. *Behav. Res. Ther.,* 1963, in press.

Ayllon, T., & Haughton, E. Control of the behavior of schizophrenic patients by food. *J. exp. Anal. Behav.,* 1962, **5,** 343–352.

Ayllon, T., & Michael, J. The psychiatric nurse as a behavioral engineer. *J. exp. Anal. Behav.,* 1959, **2,** 323–334.

Bachrach, A. J., Erwin, W., & Mohr, J. P. The control of eating behavior in an ano-

rexic by operant conditioning techniques. In L. Ullman & L. Krasner (Eds.), *Case studies in behavior modification.* New York: Holt, Rinehart, & Winston, in press.

BAER, D. M. Laboratory control of thumb-sucking by withdrawal and re-presentation of reinforcement. *J. exp. Anal. Behav.*, 1962, **5**, 525–528.

BANDURA, A. Psychotherapy as a learning process. *Psychol. Bull.*, 1961, **58**, 143–157.

BANGS, J. L., & FREIDINGER, A. Diagnosis and treatment of a case of hysterical aphonia in a thirteen year old girl. *J. speech hear. Disord.*, 1949, **14**, 312–317.

BANGS, J. L., & FREIDINGER, A. A case of hysterical dysphonia in an adult. *J. speech hear. Disord.*, 1950, **15**, 316–323.

BARRETT, BEATRICE. Reduction in rate of multiple tics by free operant conditioning methods. *J. nerv. ment. Dis.*, 1962, **135**, 187–195.

BEECH, H. R. The symptomatic treatment of writer's cramp. In H. J. Eysenck (Ed.), *Behaviour therapy and the neuroses.* New York: Pergamon Press, 1960. Pp. 349–376.

BEVAN, J. R. Learning theory applied to the treatment of a patient with obsessional ruminations. In H. J. Eysenck (Ed.), *Behaviour therapy and the neuroses.* New York: Pergamon Press, 1960. Pp. 165–169.

BLOODSTEIN, O. Conditions under which stuttering is reduced or absent: A review of literature. *J. speech hear. Disord.*, 1949, **14**, 295–302.

BOND, I. K., & HUTCHISON, H. C. Application of reciprocal inhibition therapy to exhibitionism. *Canad. Med. Ass. J.*, 1960, **83**, 23–25.

BOOKBINDER, L. J. Simple conditioning vs. the dynamic approach to symptoms and symptom substitution: A reply to Yates. *Psychol. Rep.*, 1962, **10**, 71–77.

BRADY, J. P., & LIND, D. L. Experimental analysis of hysterical blindness. *AMA Arch. gen. Psychiat.*, 1961, **4**, 331–339.

CARLSON, A. J. The conditioned reflex therapy of alcohol addiction. *Quart. J. Stud. Alcohol*, 1944, **5**, 212–215.

CHERRY, C., & SAYERS, B. Experiments upon the total inhibition of stammering by external control, and some clinical results. *J. psychosom. Res.*, 1956, **1**, 233–246.

CHERRY, C., SAYERS, B., & MARLAND, P. M. Experiments on the complete suppression of stammering. *Nature, London*, 1955, **176**, 874–875.

CLARK, D. F. Fetishism treated by negative conditioning. *Brit. J. Psychiat.*, 1963, **109**, 404–407.

COHEN, L. H., HILGARD, E. R., & WENDT, G. R. Sensitivity to light in a case of hysterical blindness studied by reinforcement-inhibition and conditioning methods. *Yale J. Biol. Med.*, 1933, **6**, 61–67.

COWDEN, R. C., & FORD, L. I. Systematic desensitization with phobic schizophrenics. *Amer. J. Psychiat.*, 1962, **119**, 241–245.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

DUNLAP, K. A revision of the fundamental law of habit formation. *Science*, 1928, **67**, 360–362.

DUNLAP, K. Repetition in the breaking of habits. *Scient. Mon.*, 1930, **30**, 66–70.

DUNLAP, K. *Habits: Their making and unmaking.* New York: Liveright, 1932.

EFRON, R. The conditioned inhibition of uncinate fits. *Brain*, 1957, **80**, 251–262.

EYSENCK, H. J. The effects of psychotherapy: An evaluation. *J. consult. Psychol.*, 1952, **16**, 319–324.

EYSENCK, H. J. *The dynamics of anxiety and hysteria.* New York: Praeger, 1957.

EYSENCK, H. J. (Ed.) *Behaviour therapy and the neuroses.* New York: Pergamon Press, 1960. (a)

EYSENCK, H. J. Personality and behavior therapy. *Proc. Roy. Soc. Med.*, 1960, **53**, 504–508. (b)

EYSENCK, H. J. The effects of psychotherapy. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology.* New York: Basic Books, 1961. Pp. 697–725.

EYSENCK, H. J. Behaviour therapy, extinction, and relapse in neurosis. *Brit. J. Psychol.*, 1963, **109**, 12–18.

FERSTER, C. B. Positive reinforcement and behavioral deficits of autistic children. *Child Develpm.*, 1961, **32**, 437–456.

FERSTER, C. B., & DEMYER, M. K. The development of performance in autistic children in an automatically controlled environment. *J. chron. Dis.*, 1961, **13**, 312–345.

FERSTER, C. B., & DEMYER, M. K. A method for the experimental analysis of the behavior of autistic children. *Amer. J. Orthopsychiat.*, 1962, **32**, 89–98.

FLANAGAN, B., GOLDIAMOND, I., & AZRIN, N. Operant stuttering: The control of stuttering behavior through response-contingent consequences. *J. exp. Anal. Behav.*, 1958, **1**, 173–178.

FRANK, J. D. *Persuasion and healing.* Baltimore, Md.: Johns Hopkins Press, 1961.

FRANKS, C. M. Alcohol, alcoholism, and conditioning: A review of the literature and

some theoretical considerations. *J. ment. Sci.*, 1958, **104**, 14–33.

FREEMAN, H. L., & KENDRICK, D. C. A case of cat phobia. *Brit. med. J.*, 1960, **2**, 497–502.

FREUND, K. Some problems in the treatment of homosexuality. In H. J. Eysenck (Ed.), *Behaviour therapy and the neuroses*. New York: Pergamon Press, 1960. Pp. 312–326.

FULLER, P. R. Operant conditioning of a vegetative human organism. *Amer. J. Psychol.*, 1949, **62**, 587–590.

GLYNN, J. D., & HARPER, P. Behaviour therapy in transvestism. *Lancet*, 1961, **1**, 619.

GOLDIAMOND, I. The maintenance of ongoing fluent verbal behavior and stuttering. *J. Mathetics*, 1962, **1**, 57–95.

GORDON, H. L. Fifty shock therapy theories. *Milit. Surgeon*, 1948, **103**, 397–401.

GREENSPOON, J. Verbal conditioning and clinical psychology. In A. J. Bachrach (Ed.), *Experimental foundations of clinical psychology*. New York: Basic Books, 1962.

GROSSBERG, J. M. Behavior therapy in a case of speech phobia ("stage fright"). Technical Report No. 3, 1963, Grant RD-892P-63, United States Department of Health, Education, and Welfare, Vocational Rehabilitation Administration.

HAUGEN, G. B., DIXON, H. H., & DICKEL, H. A. *A therapy for anxiety tension reactions*. New York: Macmillan, 1958.

HILGARD, E. R., & MARQUIS, D. M. *Conditioning and learning*. New York: Appleton-Century, 1940. Pp. 297–298.

HOBBS, N. Sources of gain in psychotherapy. *Amer. Psychologist*, 1962, **17**, 741–747.

HOLMES, FLORENCE B. An experimental investigation of a method of overcoming children's fears. *Child Develpm.*, 1936, **7**, 6–30.

ISAACS, W., THOMAS, J., & GOLDIAMOND, I. Application of operant conditioning to reinstate verbal behavior in psychotics. *J. speech hear. Disord.*, 1960, **25**, 8–12.

JONES, E. L. Explorations of experimental extinction and spontaneous recovery in stuttering. In W. Johnson (Ed.), *Stuttering in children and adults*. Minneapolis: Univer. Minnesota Press, 1955. Pp. 226–231.

JONES, H. G. The application of conditioning and learning techniques to the treatment of a psychiatric patient. *J. abnorm. soc. Psychol.*, 1956, **52**, 414–420.

JONES, H. G. The behavioral treatment of enuresis nocturna. In H. J. Eysenck (Ed.), *Behaviour therapy and the neuroses*. New York: Pergamon Press, 1960. Pp. 377–403. (a)

JONES, H. G. Continuation of Yates' treatment of a tiqueur. In H. J. Eysenck (Ed.), *Behaviour therapy and the neuroses*. New York: Pergamon Press, 1960. Pp. 250–258. (b)

JONES, MARY C. The elimination of children's fears. *J. exp. Psychol.*, 1924, **7**, 382–390. (a)

JONES, MARY C. A laboratory study of fear: The case of Peter. *Pedag. Seminary*, 1924, **31**, 308–315. (b)

KALINOWSKY, L. B., & HOCH, P. B. *Somatic treatments in psychiatry*. New York: Grune & Stratton, 1961.

KANFER, F. H. Comments on learning in psychotherapy. *Psychol. Rep.*, 1961, **9**(Monogr. Suppl. No. 6), 681–699.

KANTOR, J. R. *The logic of modern science*. Bloomington, Ind.: Principia Press, 1953.

KING, G. F., ARMITAGE, S. G., & TILTON, J. R. A therapeutic approach to schizophrenics of extreme pathology: An operant-interpersonal method. *J. abnorm. soc. Psychol.*, 1960, **61**, 276–286.

KRASNER, L. The use of generalized reinforcers in psychotherapy research. *Psychol. Rep.*, 1955, **1**, 19–25.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

KUEHNER, G. F. Hypnosis in dentistry. In R. M. Dorcus (Ed.), *Hypnosis and its therapeutic applications*. New York: McGraw-Hill, 1956.

LANG, P. J., & LAZOVIK, A. D. Experimental desensitization of a phobia. *J. abnorm. soc. Psychol.*, 1963, **66**, 519–525.

LAVIN, N., THORPE, J. G., BARKER, J., BLAKEMORE, C., & CONWAY, C. Behavior therapy in a case of transvestism. *J. nerv. ment. Dis.*, 1961, **133**, 346–353.

LAZARUS, A. A. New methods in psychotherapy: A case study. *S. Afr. med. J.*, 1958, **33**, 660–663.

LAZARUS, A. A. The elimination of children's phobias by deconditioning. *Med. Proc. S. Afr.*, 1959.

LAZARUS, A. A. Group therapy of phobic disorders by systematic desensitization. *J. abnorm. soc. Psychol.*, 1961, **63**, 504–510.

LAZARUS, A. A. The treatment of chronic frigidity by systematic desensitization. *J. nerv. ment. Dis.*, 1963, **136**, 272–278.

LAZARUS, A. A., & ABRAMOVITZ, A. The use of "emotive imagery" in the treatment of children's phobias. *J. ment. Sci.*, 1962, **108**, 191–195.

LAZARUS, A. A., & RACHMAN, S. The use of systematic desensitization in psychotherapy. *S. Afr. med. J.*, 1957, **31**, 934–937.

LEHNER, G. F. J. Negative practice as a psychotherapeutic technique. *J. gen. Psychol.*, 1954, **51**, 69–82.

LEVITT, E. E. The results of psychotherapy with children: An evaluation. *J. consult. Psychol.*, 1957, **21**, 189–196.

LINDSLEY, O. R. Operant conditioning methods applied to research in chronic schizophrenia. *Psychiat. res. Rep.*, 1956, **5**, 118–139.

LINDSLEY, O. R. Characteristics of the behavior of chronic psychotics as revealed by free operant conditioning methods. *Dis. nerv. Syst. monogr. Suppl.*, 1960, **21**, 66–78.

LIVERSEDGE, L. A., & SYLVESTER, J. D. Conditioning techniques in the treatment of writer's cramp. *Lancet*, 1955, **2**, 1147–1149.

MCLAUGHLIN, F. L., & MILLAR, W. M. Employment of air-raid noises in psychotherapy. *Brit. med. J.*, 1941, **2**, 158–159.

MALLESON, N. Panic and phobia. *Lancet*, 1959, **1**, 225–227.

MALMO, R. B., DAVIS, J. F., & BARZA, S. Total hysterical deafness: An experimental case study. *J. Pers.*, 1952, **21**, 188–204.

MAX, L. W. Breaking up a homosexual fixation by the conditioned reaction technique: A case study. *Psychol. Bull.*, 1935, **32**, 734. (Abstract)

METZNER, R. Learning theory and the therapy of neurosis. *Brit. J. Psychol.*, 1961, Monogr. Suppl. No. 33.

MEYER, V. The treatment of two phobic patients on the basis of learning principles. *J. abnorm. soc. Psychol.*, 1957, **55**, 261–266.

MEYER, V., & GELDER, M. G. Behaviour therapy and phobic disorders. *Brit. J. Psychiat.*, 1963, **109**, 19–28.

MICHAEL, J., & MEYERSON, L. A behavioral approach to counseling and guidance. *Harv. educ. Rev.*, 1962, **32**, 382–402.

MOWRER, O. H. Apparatus for the study and treatment of enuresis. *Amer. J. Psychol.*, 1938, **51**, 163–165.

OSWALD, I. Induction of illusory and hallucinatory voices with considerations of behaviour therapy. *J. ment. Sci.*, 1962, **108**, 196–212.

PAVLOV, I. P. *Lectures on conditioned reflexes. Vol. 2. Conditioned reflexes and psychiatry.* (Trans. & ed. by W. H. Gantt) New York: International Univer. Press, 1941.

PETERS, H. N., & JENKINS, R. L. Improvement of chronic schizophrenic patients with guided problem-solving motivated by hunger. *Psychiat. Quart. Suppl.*, 1954, **28**, 84–101.

RACHMAN, S. The treatment of anxiety and phobic reactions by systematic desensitization psychotherapy. *J. abnorm. soc. Psychol.*, 1959, **58**, 259–263.

RACHMAN, S. Sexual disorders and behavior therapy. *Amer. J. Psychiat.*, 1961, **118**, 235–240.

RAYMOND, M. J. Case of fetishism treated by aversion therapy. *Brit. med. J.*, 1956, **2**, 854–856.

REIDER, N. The demonology of modern psychiatry. *Amer. J. Psychiat.*, 1955, **111**, 851–856.

RICKARD, H. C., DIGNAN, P. J., & HORNER, R. F. Verbal manipulation in a psychotherapeutic relationship. *J. clin. Psychol.*, 1960, **16**, 364–367.

RICKARD, H. C., & DINOFF, M. A follow-up note on "Verbal manipulation in a psychotherapeutic relationship." *Psychol. Rep.*, 1962, **11**, 506.

ROSENZWEIG, S. A transvaluation of psychotherapy: A reply to Hans Eysenck. *J. abnorm. soc. Psychol.*, 1954, **49**, 298–304.

ROTTER, J. B. *Social learning and clinical psychology*. Englewood Cliffs, N. J.: Prentice-Hall, 1954.

ROTTER, J. B. Psychotherapy. *Annu. Rev. Psychol.*, 1960, **11**, 381–414.

RUBINSTEIN, C. The treatment of morphine addiction in tuberculosis by Pawlow's conditioning method. *Amer. Rev. Tuberc.*, 1931, **24**, 682–685.

RUDOLF, G. Deconditioning and time therapy. *J. ment. Sci.*, 1961, **107**, 1097–1101.

SAUL, L. J., ROME, H., & LEUSER, E. Desensitization of combat fatigue patients. *Amer. J. Psychiat.*, 1946, **102**, 476–478.

SCHWARTZ, L. A. Group psychotherapy in the war neuroses. *Amer. J. Psychiat.*, 1945, **101**, 498–500.

SEARS, R. R., & COHEN, L. H. Hysterical anesthesia, analgesia, and astereognosis. *AMA Arch. Neurol. Psychiat.*, 1933, **29**, 260–271.

SHANE, MARY LOU. Effect on stuttering of alteration in auditory feedback. In W. Johnson (Ed.), *Stuttering in children and adults*. Minneapolis: Univer. Minnesota Press, 1955.

SHEEHAN, J. G. The modification of stuttering through non-reinforcement. *J. abnorm. soc. Psychol.*, 1951, **46**, 51–63.

SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.

SLACK, C. W. Experimenter-subject psychotherapy: A new method of introducing intensive office treatment for unreachable cases. *Ment. Hyg., N. Y.*, 1960, **44**, 238–256.

STEVENSON, I., & WOLPE, J. Recovery from sexual deviations through overcoming non-

sexual neurotic responses. *Amer. J. Psychiat.*, 1960, **116**, 737–742.

SYLVESTER, J. D., & LIVERSEDGE, L. A. Conditioning and the occupational cramps. In H. J. Eysenck (Ed.), *Behaviour therapy and the neuroses.* New York: Pergamon Press, 1960.

TERHUNE, W. B. The phobic syndrome. *AMA Arch. Neurol. Psychiat.*, 1949, **62**, 162–172.

THOMPSON, G., & BIELINSKI, B. Improvement in psychosis following conditioned-reflex treatment for alcoholism. *J. nerv. ment. Dis.*, 1953, **117**, 537–543.

WALTON, D. The application of learning theory to the treatment of a case of bronchial asthma. In H. J. Eysenck (Ed.), *Behaviour therapy and the neuroses.* New York: Pergamon Press, 1960. Pp. 188–189. (a)

WALTON, D. The application of learning theory to the treatment of a case of neurodermatitis. In H. J. Eysenck (Ed.), *Behaviour therapy and the neuroses.* New York: Pergamon Press, 1960. Pp. 272–274. (b)

WALTON, D. The relevance of learning theory to the treatment of an obsessive-compulsive state. In H. J. Eysenck (Ed.), *Behaviour therapy and the neuroses.* New York: Pergamon Press, 1960. Pp. 153–164. (c)

WALTON, D. Strengthening of incompatible reactions and the treatment of a phobic state in a schizophrenic patient. In H. J. Eysenck (Ed.), *Behaviour therapy and the neuroses.* New York: Pergamon Press, 1960. Pp. 170–180. (d)

WALTON, D. Experimental psychology and the treatment of a ticqueur. *J. child Psychol. Psychiat.*, 1961, 2, 148–155.

WALTON, D., & BLACK, D. A. The application of learning theory to the treatment of stammering. *J. psychosom. Res.*, 1958, **3**, 170–179.

WALTON, D., & BLACK, D. A. The application of modern learning theory to the treatment of chronic hysterical aphonia. In H. J. Eysenck (Ed.), *Behaviour therapy and the neuroses.* New York: Pergamon Press, 1960. Pp. 259–271.

WEINBERG, N. H., & ZASLOVE, M. "Resistance" to systematic desensitization of phobias. *J. clin. Psychol.*, 1963, **19**, 179–181.

WHITE, J. G. The use of learning theory in the psychological treatment of children. *J. clin. Psychol.*, 1959, **15**, 227–229.

WICKLAND, R. F. A new approach to the treatment and rehabilitation of hospitalized chronic schizophrenics. Paper read at Western Psychological Association, Santa Monica, May 1963. (Mimeo)

WILLIAMS, C. D. The elimination of tantrum behavior by extinction procedures. *J. abnorm. soc. Psychol.*, 1959, **59**, 269.

WISCHNER, G. J. Stuttering behavior and learning: A preliminary theoretical formulation. *J. speech hear. Disord.*, 1950, **15**, 324–335.

WOLF, M., MEES, H., & RISLEY, T. Application of operant conditioning procedures to the behavior problems of an autistic child. Paper read at Western Psychological Association, Santa Monica, May 1963. (Mimeo)

WOLPE, J. *Psychotherapy by reciprocal inhibition.* Stanford: Stanford Univer. Press, 1958.

WOLPE, J. Reciprocal inhibition as the main basis of psychotherapeutic effects. In H. J. Eysenck (Ed.), *Behaviour therapy and the neuroses.* New York: Pergamon Press, 1960. Pp. 88–113.

WOLPE, J. The systematic desensitization treatment of neuroses. *J. nerv. ment. Dis.*, 1961, **132**, 189–203.

WOLPE, J. The experimental foundations of some new psychotherapeutic methods. In A. J. Bachrach (Ed.), *Experimental foundations of clinical psychology.* New York: Basic Books, 1962. Pp. 554–575. (a)

WOLPE, J. Isolation of a conditioning procedure as the crucial psychotherapeutic factor: A case study. *J. nerv. ment. Dis.*, 1962, **134**, 316–329. (b)

YATES, A. J. The application of learning theory to the treatment of tics. *J. abnorm. soc. Psychol.*, 1958, **56**, 175–182. (a)

YATES, A. J. Symptoms and symptom substitution. *Psychol. Rev.*, 1958, **65**, 371–374. (b)

ZAX, M., & KLEIN, A. Measurement of personality and behavior changes following psychotherapy. *Psychol. Bull.*, 1960, **57**, 435–448.

ZIMMERMAN, ELAINE H., & ZIMMERMAN, J. The alteration of behavior in a special classroom situation. *J. exp. Anal. Behav.*, 1962, **5**, 59–60.

(Received September 16, 1963)

Psychological Bulletin
1964, Vol. 62, No. 2, 89-109

FUNCTIONS OF THE AMYGDALA [1]

GRAHAM V. GODDARD [2]

McGill University

Anatomically the amygdala is part of the limbic system. It receives input from all sensory modalities and from interpretive cortex and it sends impulses to basic subcortical structures such as the hypothalamus and reticular formation. The evidence shows that the amygdala is very important in the regulation of drives, especially fear, but is not itself essential for their elaboration. It appears that the amygdala is primarily involved in the suppression of motivated approach behavior and in the association of fear and avoidance behavior with previously neutral stimuli.

In man, the amygdaloid complex or amygdala is a group of closely associated nuclei situated deep in the temporal lobe. Lying partly beneath the uncus, and beside the hippocampus, it becomes blended with the grey matter on its inferomedial aspect, and with the lentiform nucleus dorsally. The differentiation of the various nuclei within the amygdala has remained essentially unchanged from marsupials to man. The study of more primitive animals, however, shows a phylogenetic division into a primitive centromedial group and a more recent basolateral group (Johnson, 1923; Koikegami, Fuse, Yokoyama, Watanabe, & Watanabe, 1955).

The centromedial group is composed of the central and medial nuclei and the nucleus of the lateral olfactory tract. The small-celled medial portion of the basal nucleus is also sometimes included. Many authors consider the cortical nucleus to belong to the centromedial group, and it is even common to use the name "corticomedial group" (Gloor, 1960; Johnson, 1923; Kaada, Andersen, & Jensen, 1954; Pribram & Kruger, 1954). However, there is a lack of agreement about this (Koikegami et al., 1955), so it is probably wiser to leave the cortical nucleus unclassified.

The basolateral group includes the lateral nucleus and the large-celled lateral portion of the basal nucleus. It is also usual to include the medial portion of the basal nucleus, but again, this is not unanimous (Koikegami et al., 1955).

The afferent connections to the amygdala are not very well defined. There is a considerable olfactory input from the lateral olfactory tract into the cortical and centromedial nuclei, and there are direct connections from the piriform cortex (Fox, 1940; Gloor, 1960; Johnson, 1923; Kaada, 1951; Pribram, Lennox, & Dunsmore, 1950), the inferior temporal gyrus (Whitlock & Nauta, 1956), and probably from the temporal pole (Klingler & Gloor, 1960; Pribram et al., 1950; Segundo, Naquet, & Arana, 1955). The amygdala also receives impulses from the brain-stem reticular formation (Machne & Segundo, 1956), the cerebellum (Anand, Malhotra, Singh, & Dua, 1959a), the hippocampus (Gloor, 1960), and the thalamus (Wendt & Albe-Fessard, 1962). The stria terminalis, which in the past has been considered only as an efferent pathway from the amygdala, has recently been shown to carry important

[1] Part of a doctoral thesis written under the supervision of P. M. Milner, McGill University. Supported by a National Research Council of Canada Studentship.

[2] Now at the University of Waterloo, Ontario.

afferent connections to the amygdala (Hilton & Zbrozyna, 1963; Zbrozyna, 1960).

The amygdala also receives input from all the various sensory modalities (Bonvallet, Dell, & Hugelin, 1952; Gloor, 1960). Microelectrode recordings have demonstrated that there is often convergence from many modalities on to the same amygdaloid cell (Machne & Segundo, 1956; Sawa & Delgado, 1963). "Attention attracting" and meaningful stimuli appear to have stronger effects and Wendt and Albe-Fessard (1962) have demonstrated that the somatic stimuli are mediated via the ventral posterolateral nucleus of the thalamus and the somatic cortical Area II.

The efferent connections of the amygdala are extremely widespread. Gloor (1960) has recently published a review article in which he thoroughly discusses the amygdaloid projection system. There has been no recent evidence that significantly modifies this treatise. The primary projection field, characterized by short latency monosynaptic connections, consists of the basal septal region, head of the caudate nucleus, preoptic area, anterior hypothalamus, ventromedial hypothalamus, contralateral amygdala, piriform cortex, anterior temporal cortex, and insular cortex. The secondary projection field, characterized by longer latency multisynaptic connections consists of the remainder of the hypothalamus, subthalamus, entopeduncular nucleus, mesencephalic tegmentum, hippocampus, diffuse thalamic projection system, and ultimately most of the neocortex (Gloor, 1960). To this should be added the nuclei medialis dorsalis, lateralis posterior, and lateralis dorsalis of the thalamus (Fox, 1949) which are the thalamic projections to associative cortex (Pribram, 1960).

It can be seen, therefore, that the amygdala is not only strongly connected to the limbic "emotion" system, but has a diffuse projection, not unlike the reticular system (Adey, 1958; Feindel & Gloor, 1954), and does in fact modify the activity of the reticular system (Adey, Buchwald, & Lindsley, 1960). It is not surprising that stimulation or ablation of the amygdala has been found to affect many types of behavior. Effects that have been produced by experimental intervention with amygdaloid function include changes in eating, drinking, sexual and maternal behavior, fear, aggression, awareness, memory, as well as many autonomic and endocrine changes.

General Arousal

Following bilateral removal of the amygdala, animals are frequently depressed. Body temperature may be reduced for several days (Anand & Brobeck, 1952) and the animals may remain lethargic and sphinxlike for a week or more. During this time they do not clean themselves and frequently require forced feeding (Koikegami et al., 1955; Schreiner & Kling, 1953b; Yamada & Greer, 1960).

This lethargy gradually wears off, and the animal may become hyperactive (Schreiner & Kling, 1953b; Turner, 1954; Wood, 1958; Wood, 1956). The hyperactivity appears to be based on an increased reactiveness to environmental stimuli, particularly visual stimuli. The animals examine and reexamine everything around them. This examination involves a great deal of sniffing and licking (Brady, Schreiner, Geller, & Kling, 1954; Koikegami, Dodo, Mochida, & Takahashi, 1957; Schreiner & Kling, 1954).

Electrical stimulation of the amygdala causes restlessness (Gastaut, Vigouroux, Corriol, & Badier, 1951) and

has been shown to produce cortical desynchronization in the cat, dog, monkey, and human (Feindel, 1961; Kaada, 1951). This EEG activation is much like that produced by stimulation of the reticular formation (Feindel & Gloor, 1954), and may be part of the orientation response which has been shown to be mediated by almost all the amygdaloid projection fibers (Kaada & Ursin, 1957b; Ursin & Kaada, 1960b).

Dietary Changes

Bilateral ablation of the amygdala has been shown to affect food intake in the rat, cat, dog, monkey, and human. There is little agreement, however, on the exact nature of the change. While a few authors find no difference at all (Anand & Brobeck, 1952; Wood, 1956), some find a striking decrease in food intake (Alajouanine, Villey, Nehlil, & Houdart, 1957; Anand, Dua, & Chhina, 1957a; Brutkowski, Fonberg, Kreiner, Mempel, & Sychowa, 1962; Brady et al., 1954; Kling, Orbach, Schwartz, & Towne, 1960; Kling & Schwartz, 1961a, 1961b; Schreiner & Kling, 1953b; Weiskrantz, 1956; Yamada & Greer, 1960) which may last for a few days or as long as several weeks. Other authors report an increased food intake (Alajouanine et al., 1957; Fuller, Rosvold, & Pribram, 1957; Green, Clemente, & DeGroot, 1957; Koikegami, Fuse, Hiroki, Kazami, & Kageyama, 1958; Morgane & Kosman, 1957, 1959, 1960; Schwartzbaum, 1961; Wood, 1958) which gradually develops and then may either slowly disappear or lead to a permanent obesity. In cases where both a decrease and an increase in eating are found in the same animal, the decrease always precedes the increase. It is possible therefore that the general depression discussed above accounts for the decreased food intake.

The increased food intake may be accounted for by a constant nibbling where the animal takes all small objects into its mouth, ejecting only those which are not edible. This oral behavior, however, does not always lead to an increase in body weight (Klüver & Bucy, 1938; Pribram & Bagshaw, 1953; Rosvold, Fuller, & Pribram, 1951; Sawa, Ueki, Arita, & Harada, 1954; Schreiner & Kling, 1953a).

Amygdaloid hyperphagia does not seem to result from increased food motivation. Masserman, Levitt, McAvoy, Kling, and Pechtel (1958) claim that amygdalectomy in cats and adult monkeys decreases food motivation. Schwartzbaum (1960a, 1961) has shown that amygdalectomized monkeys which became hyperphagic on ad lib feeding can be trained to press a bar for food but are relatively unresponsive to variations in length of food deprivation and differing reward characteristics.

Perhaps the most consistent picture to emerge is that bilateral amygdaloid removal causes a severe but transient decrease in food intake, followed by a longer lasting and sometimes permanent hyperphagia. This hyperphagia is characterized by constant nibbling and does not involve high food motivation.

In view of the many experiments showing that amygdaloid lesions affect eating, there have been surprisingly few stimulation studies on this problem. Fonberg and Delgado (1961) have shown that stimulation of the amygdala in cats immediately inhibits eating and the effect may last for several hours or even days after the experience. They also found that the long-term effects were specific to the type of food being eaten at the time of stimulation. On the other hand Robinson and Mishkin (1961), in a preliminary report, claim that amygdaloid stimulation in monkeys will cause food and water inges-

tion. The number of animals studied in both these experiments is very small and any interpretation of the results would be very tentative.

The role of the amygdala in drinking has not been studied, although the animals that do not eat following amygdaloid ablation probably do not drink much either (Brutkowski et al., 1962; Yamada & Greer, 1960).

Sexual Behavior

Sexual behavior in the rat, cat, lynx, agouti, dog, monkey, and human has been altered by amygdaloid lesions. Most authors have found that a few weeks after bilateral ablation of the amygdala many animals, particularly males, become hypersexual (Anand, Chhina, & Dua, 1959; Brady et al., 1954; Fuller et al., 1957; Green et al., 1957; Kling et al., 1960; Masserman et al., 1958; Orbach, Milner, & Rasmussen, 1960; Pribram & Bagshaw, 1953; Sawa et al., 1954; Schreiner & Kling, 1953a, 1953b, 1954, 1956; Terzian & Ore, 1955; Wood, 1958). Not all authors have found a change (Anand et al., 1959; Green et al., 1957; Kling & Schwartz, 1961a; Martin, Endröczi, & Bata, 1958; Morgane & Kosman, 1959; Orbach et al., 1960; Weiskrantz, 1956) and only some of these results can be accounted for by an insufficiently long postoperative observation period.

Although the hypersexuality involves an increase in the amount of sexual behavior, its most striking characteristic is that it involves an increase in the diversity of sexual behavior. The animal will attempt to copulate with nonestrous females, males, members of other species, and even pieces of cloth. The hypersexuality is abolished by castration, by lesions of the ventromedial nucleus of the hypothalamus (Schreiner & Kling, 1954), and by septal lesions (Kling et al., 1960). In man the hypersexuality is rather mild, consisting of holding hands (Sawa et al., 1954) or exhibitionism and masturbation (Terzian & Ore, 1955).

Many of the above-mentioned studies included removal of the piriform cortex, and Green et al. (1957) found that the piriform removals alone account for the sexual changes. Undoubtedly the piriform cortex plays a major role, but it cannot be considered to account for the entire effect as Wood (1958) has shown that lesions restricted to the lateral nucleus of the amygdala will, after a delay of several weeks, cause hypersexuality.

Several authors have found that amygdalectomy causes a decrease in sexual behavior (Kling & Schwartz, 1961a; Thompson & Walker, 1950, 1951; Walker, Thompson, & McQueen, 1953). In all cases this decrease was only during the first week or two after the operation when the animals groom little and frequently do not eat.

Paradoxically, although amygdaloid lesions cause more hypersexuality in males than in females, the effects on the genital organs show the opposite picture. Amygdalectomy in adult male rats and cats causes a significant degeneration of the testes, whereas in the adult female cat the ovaries remain unaffected (Greer & Yamada, 1959; Kling et al., 1960; Yamada & Greer, 1960); and in the young female cat, amygdalectomy causes a precocious development of ovaries, uterus, and vagina (Elwers & Critchlow, 1961; Lundberg, 1962). Furthermore, electrical stimulation of the amygdala in female rats, rabbits, cats, and dogs produces ovulation and increased uterine movements (Bunn & Everett, 1957; Koikegami, Yamada, & Usui, 1953; Shealy & Peele, 1957). The major effect on female behavior following amygdaloid lesions

seems to be disruption of maternal behavior. In cases where an amygdalectomized monkey has given birth the infant has a very limited chance of survival (Masserman et al., 1958; Walker et al., 1953).

There has been little work done on the effects of amygdaloid stimulation on sexual behavior, no doubt because of the problems of electrode leads becoming entangled. However, Lissák and Endröczi (1961) have reported sexual hyperactivity in male and female cats and dogs during the poststimulation period. Alonso-deFlorida and Delgado (1958) stimulated cats in a group situation but did not report any sexual behavior other than males which crouched in a manner analogous to estrous females.

Attempts to measure the strength of the sex drive in amygdaloid hypersexual animals have not been made. However the effects on the gonads of these animals would suggest that, as in the case of amygdaloid hyperphagia, the drive is not very strong. It has been reported (Schreiner & Kling, 1954) that a hypersexual male will attempt to mount a nonestrous female even though he gets scratched and bitten. But these cats were placid and unresponsive to noxious stimuli at any time.

Fear, Rage, and Arousal

Emotionality and responsiveness to noxious stimuli are the most controversial and frequently studied phenomena to be linked with amygdaloid function. The lesion studies which bear on this problem most frequently report placidity but sometimes report savageness. The stimulation studies most frequently report fear or rage and only occasionally a decreased aggression. Similarly the studies on hormone changes usually, although not always, support the hypothesis that the amygdala is important in the development of emotions related to fear. A further source of evidence is psychomotor epilepsy in which the seizure activity originates in or near the amygdala. The seizures are accompanied by fear states far more frequently than epileptic seizures originating from other areas of the brain.

In 1938 Klüver and Bucy attracted considerable attention by elaborating an experiment first performed by Brown and Schäfer in 1888. Bilateral removal of the temporal lobes in monkeys caused inability to recognize objects visually, oral exploration of the environment, lack of emotion, and continual curiosity for objects in the immediate environment. The visual changes have since been shown to depend on the inferior temporal cortex (Mishkin, 1954; Mishkin & Pribram, 1954; Pribram & Bagshaw, 1953). The only exception is that amygdaloid lesions appear to affect brightness transposition (Schwartzbaum & Pribram, 1960). All other effects have been reproduced to a lesser degree by bilateral removal of the amygdala, although it should be pointed out that the cortex of the orbital gyrus, insula, temporal tip, and anterior cingulate, is also of considerable importance to these functions.

The lack of emotion, usually described as placidity, involves a loss of aggression that is particularly striking in wild animals like the lynx and the Norway rat, a reduced responsiveness to normally noxious stimuli, and a completely fearless curiosity for dangerous or threatening stimuli including members of other species. This placidity following bilateral lesions involving the amygdala has been reported for the rat, cat, lynx, agouti, dog, phalanger, monkey, and human (Adey, 1958; Anand & Brobeck, 1951; Anand et al., 1959; Anand et al., 1957a; Brady et al., 1954; Brown & Schäfer, 1888; Brutkowski,

Fonberg, & Mempel, 1961; Downer, 1962; Fuller et al., 1957; Green et al., 1957; Horvath, 1963; Kling & Hutt, 1958; Kling et al., 1960; Klüver, 1958; Klüver & Bucy, 1938; Koikegami et al., 1958; Lissák & Endröczi, 1961; Martin et al., 1958; Masserman et al., 1958; Morin, Gastaut, Vigouroux, & Roger, 1952; Orbach et al., 1960; Poirier, 1952; Pribram & Bagshaw, 1953; Pribram, Mishkin, Rosvold, & Kaplan, 1952; Pribram & Weiskrantz, 1957; Rosvold et al., 1951; Rosvold, Mirsky, & Pribram, 1954; Sawa et al., 1954; Schreiner & Kling, 1953a, 1953b, 1954, 1956; Scoville, Dunsmore, Liberson, Henry, & Pepe, 1953; Smith, 1950; Summers & Kaelber, 1962; Terzian & Ore, 1955; Thompson & Walker, 1950, 1951; Turner, 1954; Walker et al., 1953; Weiskrantz, 1953, 1956; Wood, 1956).

In some of the above 43 studies the placidity was fairly mild or did not last for more than a few weeks. The more usual finding, however, is that the placidity starts immediately after the operation and is relatively permanent. In most of the studies the lesions have included structures other than the amygdala. However a sufficient number of control lesions have been made in these other structures to make it almost certain that the placidity is primarily due to removal of the amygdala. When large amounts of the temporal lobe are removed along with the amygdala the placidity effects are usually more pronounced. Occasionally some placidity is obtained from removing the inferotemporal, piriform, and entorhinal cortex while sparing the amygdala (Adey, Merrillees, & Sunderland, 1956; Akert, Gruesen, Woolsey, & Meyer, 1961; Meyer, 1958). This latter result is not surprising as these structures supply many of the afferent connections to the amygdala.

In rats, hyperemotionality that has been produced by septal lesions (King & Meyer, 1958) or hypothalamic lesions (Anand & Brobeck, 1951) is attenuated by the addition of amygdaloid lesions. The addition of amygdaloid lesions in cats will partially relieve hypothalamic rage (Kling & Hutt, 1958), but only occasionally reduces septal rage (Kling et al., 1960). However in dogs that have been made placid by amygdaloid lesions, the addition of lesions of the rostral cingulate cortex including the septum causes them to become wild. Obviously, although the amygdala is important in rage, it is not essential. Even the placid animals may show rage if sufficiently provoked; they merely have a very high threshold.

In fact there are a number of studies showing that bilateral removal of the amygdala in cats leads to rage (Bard & Mountcastle, 1947; Bard & Rioch, 1937; Spiegel, Miller, & Oppenheimer, 1940; Wood, 1958). A number of hypotheses have been considered in an attempt to explain the discrepancy between these studies and the studies reporting placidity. However it seems that differences in species, surgical techniques, size or location of lesion, involvement of different extra-amygdaloid structures and pre- or postoperative experience cannot be considered as complete explanations. Schreiner and Kling (1953b) and Kling et al. (1960) have produced rage in a small number of cats using exactly the same technique as that which produced placidity in the other cats. Furthermore, they were unable to find any histological differences in these animals that correlated with the behavioral differences.

Green et al. (1957) also found that following amygdaloid ablation some cats were placid and others showed rage. This rage was found to develop only if the lesions involved the hippo-

campus, and then only if the animals developed periodic seizures. It has never been demonstrated, but it is not unreasonable to suggest, that in all cases where rage follows amygdalectomy it is due to a discharging focus on the periphery of the lesion. This could easily be investigated with deep recording electrodes, the only difficulty being in obtaining amygdala-lesioned savage animals. Some indirect support for this hypothesis comes from the fact that when amygdalectomy causes rage, the rage takes a long time, sometimes many weeks, to develop.

The only study to which this hypothesis probably does not apply is that of Wood (1958). Wood found that small bilateral electrolytic lesions in the cat increase aggressiveness only when placed in the basal or central nuclei of the amygdala. It is possible that subtotal lesions such as these can considerably alter the activity of the remaining amygdaloid nuclei.

One other study should be mentioned in this context. Rosvold et al. (1954) have claimed that amygdalectomy in monkeys causes a drop in social dominance (i.e., placidity) if the animals are kept in a group situation postoperatively, but an increase in aggressiveness if they are kept in individual cages. However, their test of aggression was whether the animals approached or withdrew from threatening stimuli, and, as mentioned earlier, placidity involves fearlessly approaching threatening stimuli. The authors were aware of this, and in a footnote stated that the term fearless would be a better description of the behavior than the term aggressive.

The role of the amygdala in fear and aggression has also been studied using electrical stimulation. The behavioral pattern that results from gradually increasing the intensity of amygdaloid stimulation is first for the animal to cease activity and appear to be attending to something. Then the head is turned as if the animal is searching for something; this is frequently accompanied by sniffing, chewing, swallowing, and twitching of the ipsilateral facial muscles. At higher intensities salivation and piloerection occur, and frequently clonic movements of the ipsilateral forepaw. Vocalization or hissing may occur at this point and sometimes urination and even defecation. The behavior is usually interpreted to represent fear or defense.

If the stimulation is continued or increased in intensity the animal may either suddenly take flight, or else go into a seizure involving circling, rolling over, and clonic movements of all four limbs. In other cases the animal may suddenly show rage. This rage may or may not be well directed. Fear and rage have been elicited by amygdaloid stimulation primarily in cats (Andersen, Jansen, & Kaada, 1952; Fernandez-de-Molina & Hunsperger, 1959, 1962; Gastaut, 1952; Gastaut, Naquet, Vigouroux, & Corriol, 1952; Gastaut et al., 1951; Hilton & Zbrozyna, 1963; Hunsperger, 1956, 1959; Kaada et al., 1954; Kaada, Jansen, & Andersen, 1952; Kaada & Ursin, 1957a; MacLean & Delgado, 1953; Magnus & Lammers, 1956; Sano, 1958; Shealy & Peele, 1957; Ursin & Kaada, 1960a; Wood, 1958; Zbrozyna, 1960).

Similarly in monkeys, amygdaloid stimulation can cause fear and rage (Anand & Dua, 1956d). This fear has been demonstrated to be real in that the monkey will perform a response that has previously been conditioned to the fear of receiving an electric shock (Delgado, Rosvold, & Looney, 1956). In man, although fear can sometimes be produced by amygdaloid stimulation (Brazier, Schroeder, Chapman, Geyer,

Fager, Poppen, Solomon, & Yakovlev, 1954; Chapman, Schroeder, Geyer, Brazier, Fager, Poppen, Solomon, & Yakovlev, 1954; Delgado, 1960; Feindel, 1961; King, 1961), rage is almost never encountered. Mullan and Penfield (1959), in the course of stimulating 213 patients in the amygdaloid area, encountered fear 23 times but never once observed rage. However it should be reported that Heath, Monroe, and Mickle (1955) elicited rage from one schizophrenic woman, and King (1961) has elicited statements such as "I want to hit something" and "I want to tear things."

Although rage is not obtained from all animals that are stimulated in the amygdala, there is only one study in which amygdaloid stimulation has been shown to actually inhibit aggression. Egger and Flynn (1962) found that a cat placed with a rat would attack it during hypothalamic stimulation, but that this attack could be prevented by simultaneous amygdaloid stimulation.

A finding related to the role of the amygdala in emotion is that amygdaloid stimulation frequently produces facial movements which are not mediated by the motor cortex (Andy & Mukawa, 1960; Baldwin, Frost, & Wood, 1956). It seems that the likelihood of producing these grimaces increases as one goes up the phylogenetic scale paralleling the increase in reflexive facial movements that normally accompany emotions.

As would be expected from the results on fear and aggressive behavior, the amygdala is important for control of various autonomic responses. Bilateral removal of the amygdala in cats and monkeys has been shown to affect respiration, heart rate, blood pressure, and blood-sugar level (Anand et al., 1959; Anand, Dua, & Chhina, 1957b; Kling et al., 1960). The direction of change is rather variable but the most frequent response is for respiration and heart rate to decrease and blood pressure and blood glucose to increase.

The effects of amygdaloid stimulation on autonomic responses have been more widely studied. Pupil dilation and piloerection have already been mentioned. Respiration, heart rate, and blood pressure have been studied in the rat, rabbit, cat, dog, monkey, and human (Anand & Dua, 1956b; Andy, Bonn, Chinn, & Allen, 1959; Baldwin, Frost, & Wood, 1954; Chapman et al., 1954; Gastaut, 1952; Hilton & Zbrozyna, 1963; Kaada, 1951; Koikegami et al., 1957; Koikegami, Kimoto, & Kido, 1953; MacLean & Delgado, 1953; Naquet, 1953, 1954; Turner, 1954; Ursin & Kaada, 1960a; Wood, Schottelius, Frost, & Baldwin, 1958). Here again, the exact nature of the response varies considerably. There is usually an inhibition of respiration which is followed by either an increase or decrease in rate, and either an increase or decrease in amplitude. Similarly the changes in heart rate and blood pressure may involve either an increase or a decrease. Anand and Dua (1956a) found that amygdaloid stimulation in cats and monkeys caused an increase in blood-sugar levels.

If the amygdala is important in controlling fear and aggression it is to be expected that it not only enters into the regulation of the autonomic nervous system but also acts on the endocrine system, particularly the endocrine responses to stress. Generally speaking this has been shown to be true.

The endocrine system appears to be generally depressed after removal of the amygdala in rats, dogs, and goats, but primarily if the removal is made early in life (Koikegami et al., 1958; Koikegami et al., 1955). These animals have impaired growth and atrophy of

the parotid, thyroid, pituitary adrenal, and pancreatic glands. If amygdaloid lesions are placed in adult animals (rat, rabbit, cat, and monkey), the endocrine effects are less severe. Apart from testicular degeneration the only gland which is much affected is the adrenal gland. The actual weight of the adrenal gland does not change very much except for a small increase during the first week or two postoperatively (Greer & Yamada, 1959; Kling et al., 1960; Yamada & Greer, 1960). However Martin et al. (1958) found that amygdalectomy in cats and dogs dramatically increased the level of adrenal corticoids in the blood, and sometimes a new type of corticoid was produced.

Other authors have concerned themselves with the adrenal response to stress in amygdalectomized animals. Knigge (1961) found that amygdalectomy in rats caused a delay in the corticosterone response to immobilization. Mason, Nauta, Brady, and Robinson (1959) found that the 17-hydroxy-corticosteroid response to avoidance training was abolished by total amygdaloid removals. This type of result would be expected from the ensuing placidity. However Anand et al. (1957b) report the eosinophil response to subcutaneous injection of saline is not affected by lesions including the amygdala, and Bovard and Gloor (1961) found that small lesions in the central nucleus of the amygdala increased the corticosterone response to immobilization in rats. This is the location where Wood (1958) found that small lesions increased aggressiveness so perhaps this latter finding is not really a contradiction of the other studies.

Electrical stimulation of the amygdala has been shown to cause an increase in the production of 17-hydroxy-corticosteriods. Mason (1958, 1959) has shown this in monkeys, and Mandell, Chapman, Rand, and Walter (1963) obtained similar results in humans using an intensity which was too low to produce any detectable behavioral or subjective changes.

The other major endocrine effect is for the stimulation to cause an increase in gastric secretions (Anand & Dua, 1956c; Sen & Anand, 1957; Shealy & Peele, 1957). Gastric and intestinal movements may be either increased or decreased (Eliasson, 1952; Gastaut, 1952; Koikegami, Kimoto, & Kido, 1953; Koikegami, Kushiro, & Kimoto, 1953; Shealy & Peele, 1957). Repeated amygdaloid stimulation has been shown to cause gastric bleeding and ulceration of the gastroduodenal junction (Sen & Anand, 1957). Bleeding has also been observed in the lungs (Koikegami & Fuse, 1952a, 1952b). Such changes are suggestively similar to the psychosomatic diseases that are supposed to result from psychological stress.

The amygdala, therefore, plays an important role in fear and anger, and is in a position to regulate the autonomic and endocrine components of this behavior. Further support of this statement comes from the reports of emotions associated with epilepsy. These studies have recently been reviewed by Gloor and Feindel (1963). There appears to be fairly general agreement among the various authors that about 40% of epileptic patients have emotions such as fear, anxiety, or depression as part of the ictal syndrome. Most of these patients have a focus in the vicinity of the amygdala. Ictal aggression is quite rare, but when it does occur the patient is again found to have a focus in the amygdaloid region.

Groethuysen obtained surface and depth recordings of an epileptic attack that had been triggered in an epileptic female by a structured traumatic inter-

view. The seizure activity was first picked up on the electrode near the left amygdala. This activity slowly spread to all other electrodes, until after 11 minutes the patient had a generalized convulsion (Groethuysen, Robinson, Haylett, Estes, & Johnson, 1957).

Approach and Avoidance Learning

If the amygdala is primarily involved in the regulation of fear, then it is to be expected that interference with the amygdala will disrupt avoidance learning more than approach learning. Roughly speaking, this has been found to be the case as long as the hippocampus remains intact.

Amygdalectomized dogs are able to learn an olfactory discrimination, leg flexion, and an approach response for food within normal limits (Allen, 1941; Brutkowski, Fonberg, & Mempel, 1960). If the hippocampus is also removed they have trouble relearning a visual discrimination in a T maze (Fuller et al., 1957).

Amygdalectomized monkeys are quite normal in acquiring a delayed response and various discrimination problems (Mahut & Cordeau, 1963; Orbach et al., 1960; Schwartzbaum & Pribram, 1960). If the hippocampus is also removed they have difficulty (Orbach et al., 1960).

The only outstanding contradiction of these results comes from Masserman et al. (1958) who report that amygdaloid lesions in cats and monkeys cause a decrease in attention to task and food motivation with a concomitant drop in their ability to learn new problems.

There is one thing, however, in the approach situation that amygdalectomy does disrupt. This is the withholding of a learned response. For example, if a differential conditioned stimulus is presented signaling that the animal should not respond, amygdalectomized dogs learn very slowly (Allen, 1941; Brutkowski et al., 1960). Similarly, Mahut and Cordeau (1963) have shown that amygdalectomized monkeys show considerable perseveration in delayed reversal and spatial reversal problems. This type of deficit might account for the findings of Endröczi and Lissák (1962) who showed that amygdaloid stimulation does not prevent a conditioned approach response but causes an increase in the number of intertrial responses.

Almost all the studies on avoidance learning have demonstrated a deficit of some type. Robinson (1963) has shown that amygdaloid lesions retard the rate of acquisition of a conditioned avoidance response. King (1958), using a shuttle box, found that rats with small amygdaloid lesions had a longer latency of response, but were not otherwise impaired.

Kling et al. (1960) report that although amygdaloid lesions in cats do not increase the total number of trials required to reach criterion in a shuttle box, significantly more trials are required before the animals start to make avoidances. They interpret this to mean that the cats had difficulty in learning the significance of the conditioned stimulus but not in learning to perform the conditioned response. Brady et al. (1954) have found that amygdalectomized cats were retarded in the acquisition of a conditioned avoidance but retained a preoperatively trained avoidance. Horvath (1963), also using cats, found that amygdaloid lesions disrupted both the acquisition and retention of a simple avoidance, and retarded to a lesser degree the acquisition of a simple avoidance and a passive avoidance. Using dogs, Fonberg, Brutkowski, and Mempel (1962) found that amygdalectomy disrupts retention of a

recently acquired defense response but not a highly overtrained leg-flexion avoidance.

In a series of experiments on monkeys, Weiskrantz (1956; Pribram & Weiskrantz, 1957; Weiskrantz & Wilson, 1955, 1958) has found that amygdalectomized monkeys are quick to extinguish a preoperatively acquired avoidance response in both shuttle box and panel pushing. These lesions also retarded the acquisition of both types of avoidance and a conditioned emotional response. Once having learned to make an avoidance response the animals required a stronger level of shock to make them perform.

It would seem then that bilateral removal of the amygdala in rats, cats, and monkeys will disrupt some aspect of acquiring a simple avoidance, a shuttle-box avoidance, a passive avoidance, and a conditioned emotional response. Retention of a preoperatively acquired response may or may not be disturbed.

Goddard (1963) has found similar results using a continuous low-intensity stimulation of the amygdala in rats. This stimulation, which probably causes a functional lesion, interferes with avoidance conditioning and acquisition of a conditioned emotional response but does not affect approach learning. The critical time for this stimulation to be effective is just after the unconditioned stimulus suggesting that the amygdala is important in the consolidation of the association between a neutral stimulus and a noxious event. Amygdaloid stimulation does not seem to have a very large effect on the performance of a previously established avoidance response, although Fonberg (1963) has shown that it decreases intertrial responding, and Fonberg and Delgado (1961) have shown that quite intense stimulation interfered with a wheel-turning avoidance in one out of four cats tested.

Three other studies should be mentioned in this context, although their implications are not clear due to the possible involvement of extra-amygdaloid structures. Bickford and his associates have shown that electrical stimulation of white matter 1 centimeter from the amygdala and hippocampus in man may cause a transient amnesia for events preceding the stimulation (Bickford, Mulder, Dodge, Svien, & Rome, 1956). Morrell, Roberts, and Jasper (1956) have shown that conditioned alpha blocking in monkeys is disrupted in all sensory modalities by implanting alumina cream in the amygdala. The alumina cream causes an epileptic focus which no doubt disrupts amygdaloid function, but probably also disturbs many other subcortical structures. Stamm and Knight (1963) similarly found that alumina cream placed on the medial temporal cortex near the amygdala prevents the formation of learning sets in a visual discrimination problem.

A somewhat related line of research has been conducted by Lesse (1957a, 1957b, 1960) using cats. He has recorded bursts of high amplitude 40-cycle-per-second activity in the amygdala in various behavioral situations. This activity is not seen in response to a neutral environment nor to stimuli that produce only waking, alerting, and cortical desynchronization. It is seen, however, during attack, defense, noxious stimulation, and if food is presented outside the cage of the hungry animal. Furthermore, if the cat is trained in a shuttle box the 40-cycle-per-second activity at first appears only in response to the shock, but later, as the animal starts to make avoidance responses, it appears also in response to the conditioned stimulus. If the cat

is then extinguished the 40-cycle-per-second activity stops occurring in response to the conditioned stimulus shortly after the animal stops making the conditioned response.

John and Killam (1959, 1960) also report 40-cycle-per-second activity in the amygdala during avoidance responding and shortly after the introduction of a differential stimulus during bar pressing for food. This activity ceases when the avoidance response is blocked by reserpine.

However, these studies have not controlled for respiratory changes. Domino and Ueki (1960) and Ueki and Domino (1961) have shown that similar bursts in monkeys and dogs correlate with the passage of air or other gases through the nose. Pain increases the breathing and increased breathing increases this 40-cycle-per-second activity. This is merely a correlation however. MacLean, Horwitz, and Robinson (1952) have shown that the burst activity can occur in response to pain after the insertion of a tracheal cannula which prevented olfactory input, and Gault and Leaton (1963) have shown the olfactory stimuli that cause burst activity in the olfactory bulb do not always cause activity in the amygdala. It would be very interesting to know if 40-cycles-per-second can be recorded from the amygdala of a frightened or angry animal that has been made anosmic.

Reward and Punishment

Although the amygdala has been shown to be important in avoidance learning, and amygdaloid stimulation in man causes fear and aggression much more often than pleasurable emotions, it is rather surprising that amygdaloid stimulation can often serve as the reinforcement for an instrumental response. Olds and Milner (1954) first reported that rats would press a bar to receive amygdaloid stimulation.

It has also been shown that monkeys will work to receive amygdaloid stimulation (Porter, Conrad, & Brady, 1959; Sheer, 1961). However, some electrodes in the amygdala may be positively reinforcing, while others are negatively reinforcing (Bursten & Delgado, 1958; Doty, 1961) and it has been shown that in humans the same electrode may be positive or negative depending on current intensity (Bishop, Elder, & Heath, 1963).

Wurtz and Olds (1961; Wurtz, 1962) tested each of 37 rats with amygdaloid electrodes in two situations. In one, if the rat pressed the bar, the stimulation current came on for .5 second. In the other the stimulation came on automatically every 1 second unless the bar was pressed, in which case the stimulation did not come on for 4 seconds. A response rate greater than one press every 4 seconds ensured that the stimulation remained off. Thirteen rats actively sought the stimulation, 5 rats avoided it, and 17 rats did both, despite the fact that the stimulation parameters remained the same. Only 2 rats did not press either bar. Obviously the amygdala must be classified as a reward area and also as a punishment area, but ambivalence seems to describe the most common effect.

Awareness and Interpretive Function

The discussion so far has revolved around motivational and emotional behavior; indeed these seem to be the most common findings in studies on amygdaloid function. However, investigation of epileptic discharge and amygdaloid stimulation in man has revealed changes in awareness and interpretation of the environment. This has also received some support from animal studies.

Amygdaloid stimulation in cats at fairly low intensities causes the animals to appear "attentive." It has been reported that during this state the cat is less responsive to novel and other peripheral stimuli, that is, has reduced vigilance (Gastaut, Naquet, Vigouroux, & Corriol, 1952; Gastaut, Vigouroux, & Naquet, 1952; Naquet, 1953, 1954). This was not confirmed by Ursin and Kaada (1960a), but Kaada et al. (1954) describe the cats as bewildered.

Epileptic discharge from the amygdaloid region in man frequently causes conscious confusion, sensory illusions, and feelings of unreality. This may lead to unresponsiveness to environmental stimuli and an arrest of memory recording (Feindel, 1961; Feindel & Penfield, 1954; Morris, 1956). These changes can also be produced in man by electrical stimulation of the amygdala (Feindel, 1961; Jasper & Rasmussen, 1956; Mullan & Penfield, 1959; Penfield, 1955). After bilateral amygdalectomy in schizophrenics, Sawa et al. (1954) report confusion and inability to recognize people for the first month or two postoperatively.

Topographic Localization

Gloor (1960) has prepared a table comparing the results of all the major localization studies that used amygdaloid stimulation. Since then, the only large-scale attempts have been made by Ursin and Kaada (1960a) and Fernandez-deMolina and Hunsperger (1957, 1959, 1962). The only simple statement that can be made is that there is little agreement from study to study about which nuclei yield the various behavioral changes when stimulated.

The only stimulation study reporting topographic localization that has not been contradicted is that of Wurtz (1962). Although the anatomical differentiation was very approximate, with a great deal of overlap, it seems that the centromedial group is more positively rewarding, and the basolateral group is more punishing in rat self-stimulation studies.

Another approach to the localization of function within the amygdala is to place small bilateral lesions in discrete amygdaloid nuclei. Owing to the difficulty in surgical accuracy it would probably be better to make a large ablation on one side and a small discrete lesion on the opposite side. However this has not yet been tried and the information available is all based on a few animals where small lesions have been bilaterally matched, and on larger asymmetrical lesions which have small areas in common.

Wood (1958) found that, in cats, lesions of the basal and central nuclei led to aggression, of the lateral nucleus to hypersexuality, and of the central and medial nuclei to hyperphagia. Precocious sexual development in young female rats has been attributed to medial and basal lesions by Elwers and Critchlow (1960). Stunted growth in young rats, dogs, and goats follows lesions primarily in the medial and anterior amygdala (Koikegami et al., 1958). Bovard and Gloor (1961) found that the largest effect on the corticosterone response to stress is obtained from the central nucleus.

It is impossible to say at present whether the lack of contradiction between authors using lesions is due to the limited number of studies or to a reliable topographic organization of function. The conflicting results from the various stimulation studies, however, suggest that we should be very cautious about accepting any functional localization before the results have been replicated. This is not to deny that any localization exists within the amygdala. Any such statement would be totally

unjustified. However, more refined techniques and better analyses of behavior may have to be developed before such localization is discovered.

Discussion

Anatomically the amygdala is part of the limbic system, and it is placed in a position where it can receive input from all sense modalities and from interpretive cortex. On its efferent side lie basic subcortical structures such as the hypothalamus and central grey which are important for emotional-motivational responses. It is not surprising, therefore, that the amygdala is very important in the regulation of such things as fear, but is not itself essential for their elaboration. Such considerations have led Gloor and Feindel (1963) to suggest that the amygdala is involved in correlating past experience and present sensation with the appropriate motivational responses.

A more specific hypothesis than this seems possible. It is suggested here that the amygdala is primarily involved in the active suppression of motivated approach behavior. Once an amygdalectomized animal has overcome the initial postoperative depression and lethargy, it overeats, responds sexually to all stimuli even remotely resembling a receptive female, approaches all stimuli whether dangerous or not with curiosity, and is unresponsive to variations in deprivation and food reward. In other words, it does not know when to stop. Other areas of the brain such as the hippocampus and lateral frontal cortex have been implicated in the inhibition of responses and adaptation to novel stimuli. The amygdala is involved in this system, but perhaps on a more basic level, that is, the inhibition of motivations and the satiation of drives. This may well be based on connections to the avoidance system.

Amygdalectomized animals have difficulty in associating fear and avoidance behavior with previously neutral stimuli. They also require more highly painful stimuli before they will manifest defense or aggression. When artificial activation of the amygdala is sufficiently intense to force a behavioral response this response is frequently fear, involving many of the autonomic and endocrine concomitants of pain and stress.

The positively reinforcing properties of amygdaloid stimulation are difficult to interpret, but it is probably significant that many of these points are also negatively reinforcing, and Wurtz (1962) has shown that even the purely positive points project primarily to areas of the brain which are negatively reinforcing. Furthermore, Porter et al. (1959) have shown that the activity propagated from the amygdala reward points is qualitatively different from that propagated from other reward areas.

The amygdala seems to be essential for the establishment of avoidance and the modification of approach behavior.

REFERENCES

ADEY, W. R. Organization of the rhinencephalon. In H. H. Jasper et al. (Eds.), *Reticular formation of the brain.* Boston, Mass.: Little, Brown, 1958. Pp. 621–644.

ADEY, W. R., BUCHWALD, N. A., & LINDSLEY, D. B. Amygdaloid, pallidal and peripheral influences on mesencephalic unit firing patterns with reference to mechanisms of tremor. *EEG clin. Neurophysiol.,* 1960, 12, 21–40.

ADEY, W. R., MERRILLEES, N. C. R., & SUNDERLAND, S. The entorhinal area: Behavioural evoked potential, and histological studies of its interrelationships with brain-stem regions. *Brain,* 1956, **79**, 414–439.

AKERT, K., GRUESEN, R. A., WOOLSEY, C. N., & MEYER, D. R. Klüver-Bucy syndrome in monkeys with neocortical ablations of temporal lobe. *Brain,* 1961, **84**, 480–498.

ALAJOUANINE, T., VILLEY, R., NEHLIL, J., & HOUDART, R. Troubles de l'appétit et

rhinencéphale: A propos plus particulièrement d'un cas régressif d'obésité par hyperphagie consécutif à une lobectomie temporal pour épilepsie. *Rev. neurol., Paris,* 1957, **96**, 321–323.

ALLEN, W. F. Effects of ablating the pyriform-amygdaloid areas and hippocampi on positive and negative olfactory conditioned reflexes and on conditioned olfactory differentiation. *Amer. J. Physiol.,* 1941, **132**, 81–92.

ALONSO-DEFLORIDA, F., & DELGADO, J. M. R. Lasting behavioral and EEG changes in cats induced by prolonged stimulation of amygdala. *Amer. J. Physiol.,* 1958, **193**, 223–229.

ANAND, B. K., & BROBECK, J. R. Hypothalamic control of food intake in rats. *Yale J. Biol. Med.,* 1951, **24**, 123–140.

ANAND, B. K., & BROBECK, J. R. Food intake and spontaneous activity of rats with lesions in the amygdaloid nuclei. *J. Neurophysiol.,* 1952, **15**, 421–430.

ANAND, B. K., CHHINA, G. S., & DUA, S. Effects of lesions in the limbic system on the affective behaviour and visceral responses in the monkeys and cats. *Indian J. med. Res.,* 1959, **47**, 51–58.

ANAND, B. K., & DUA, S. Blood sugar changes induced by electrical stimulation of the limbic system ("visceral brain"). *Indian J. med. Res.,* 1956, **44**, 121–124. (a)

ANAND, B. K., & DUA, S. Circulatory and respiratory changes induced by electrical stimulation of limbic system (visceral brain). *J. Neurophysiol.,* 1956, **19**, 393–400. (b)

ANAND, B. K., & DUA, S. Effect of electrical stimulation of the limbic system ("visceral brain") on gastric secretion and motility. *Indian J. med. Res.,* 1956, **44**, 125–130. (c)

ANAND, B. K., & DUA, S. Electrical stimulation of the limbic system of brain ("visceral brain") in the waking animals. *Indian J. med. Res.,* 1956, **44**, 107–119. (d)

ANAND, B. K., DUA, S., & CHHINA, G. S. Changes in the affective behaviour produced by lesions in the frontal and temporal lobes. *Indian J. med. Res.,* 1957, **45**, 353–358. (a)

ANAND, B. K., DUA, S., & CHHINA, G. S. Changes in visceral and metabolic activities after frontal and temporal lobe lesions. *Indian J. med. Res.,* 1957, **45**, 345–352. (b)

ANAND, B. K., MALHOTRA, C. L., SINGH, B., & DUA, S. Cerebellar projections to limbic system. *J. Neurophysiol.,* 1959, **22**, 451–457.

ANDERSEN, P., JANSEN, J., JR., & KAADA, B. R. Electrical stimulation of the amygdaloid nuclear complex in unanesthetized cats. *Acta psychiat. neurol. Scand., Copenhagen,* 1952, **29**, 55. (Abstract)

ANDY, O. J., BONN, P., CHINN, R. MC., & ALLEN, M. Blood pressure alterations secondary to amygdaloid and periamygdaloid after-discharges. *J. Neurophysiol.,* 1959, **22**, 51–60.

ANDY, O. J., & MUKAWA, J. Ipsilateral facial movements during amygdaloid after-discharges. *EEG clin. Neurophysiol.,* 1960, **12**, 269. (Abstract)

BALDWIN, M., FROST, L. L., & WOOD, C. D. Investigation of the primate amygdala: Movements of the face and jaws. *Neurology,* 1954, **4**, 586–598.

BALDWIN, M., FROST, L. L., & WOOD, C. D. Investigation of the primate amygdala: Movements of the face and jaws. Part 2. Effect of selective cortical ablation. *Neurology,* 1956, **6**, 288–293.

BARD, P., & MOUNTCASTLE, V. B. Some forebrain mechanisms involved in expression of rage with special reference to suppression of angry behavior. *Res. Publ. Ass. Nerv. Ment. Dis.,* 1947, **27**, 362–404.

BARD, P., & RIOCH, D. McK. A study of four cats deprived of neocortex and additional portions of the forebrain. *Johns Hopkins Bull.,* 1937, **60**, 73–148.

BICKFORD, R. C., MULDER, D. W., DODGE, H. W., JR., SVIEN, H. J., & ROME, H. P. Changes in memory function produced by electrical stimulation of the temporal lobe in man. *Res. Publ. Ass. Nerv. Ment. Dis.,* 1956, **36**, 227–243.

BISHOP, M. P., ELDER, S. T., & HEATH, R. G. Intracranial self-stimulation in man. *Science,* 1963, **140**, 394–396.

BONVALLET, M., DELL, P., & HUGELIN, A. Projections olfactives, gustatives, viscerales, vagales, visuelles, et auditives, au niveau des formations grises du cerveau antérieur du chat. *J. Physiol. Pathol. gén.,* 1952, **44**, 222–224.

BOVARD, E. W., & GLOOR, P. Effect of amygdaloid lesions on plasma corticosterone response of the albino rat to emotional stress. *Experientia,* 1961, **17**, 521–526.

BRADY, J. V., SCHREINER, L., GELLER, I., & KLING, A. Subcortical mechanisms in emotional behavior: The effect of rhinencephalic injury upon the acquisition and retention of a conditioned avoidance response in cats. *J. comp. physiol. Psychol.,* 1954, **47**, 179–186.

BRAZIER, M. A. B., SCHROEDER, H., CHAPMAN, W. P., GEYER, G., FAGER, C., POPPEN, J. L., SOLOMON, H. C., & YAKOVLEV, P. I.

Electroencephalographic recordings from depth electrodes implanted in the amygdaloid region in man. *EEG clin. Neurophysiol.,* 1954, **6**, 702.

Brown, S., & Schäfer, E. A. An investigation into the functions of the occipital and temporal lobes of the monkey's brain. *Phil. Trans. Roy. Soc. Lond. Ser. B.,* 1888, **179**, 303–327.

Brutkowski, S., Fonberg, E., Kreiner, J., Mempel, E., & Sychowa, B. Aphagia and adipsia in a dog with bilateral complete lesions of the amygdaloid complex. *Acta Biol. exp., Warsaw,* 1962, **22**, 43–50.

Brutkowski, S., Fonberg, E., & Mempel, E. Alimentary Type II (instrumental) conditioned reflexes in amygdala dogs. *Acta Biol. exp., Warsaw,* 1960, **20**, 263–271.

Brutkowski, S., Fonberg, E., & Mempel, E. Angry behaviour in dogs following bilateral lesions in the genual portion of the rostral cingulate gyrus. *Acta Biol. exp., Warsaw,* 1961, **21**, 199–205.

Bunn, J. P., & Everett, J. W. Ovulation in persistent-estrous rats after electrical stimulation of the brain. *Proc. Soc. Exp. Biol. Med.,* 1957, **96**, 369–371.

Bursten, B., & Delgado, J. M. R. Positive reinforcement induced by intracerebral stimulation in the monkey. *J. comp. physiol. Psychol.,* 1958, **51**, 6–10.

Chapman, W. P., Schroeder, H. R., Geyer, G., Brazier, M. A. B., Fager, C., Poppen, T. L., Soloman, H. C., & Yakovlev, P. I. Physiological evidence concerning importance of the amygdaloid nuclear region in the integration of circulatory function and emotion in man. *Science,* 1954, **120**, 949–950.

Delgado, J. M. R. Emotional behavior in animals and humans. *Psychiat. res. Rep.,* 1960, **12**, 259–266.

Delgado, J. M. R., Rosvold, H. E., & Looney, E. Evoking conditioned fear by electrical stimulation of subcortical structures in the monkey brain. *J. comp. physiol. Psychol.,* 1956, **49**, 373–380.

Domino, E. F., & Ueki, S. An analysis of the electrical burst phenomenon in some rhinencephalic structures of the dog and monkey. *EEG clin. Neurophysiol.,* 1960, **12**, 635–648.

Doty, R. W. Conditioned reflexes formed and evoked by brain stimulation. In D. E. Sheer (Ed.), *Electrical stimulation of the brain.* Austin: Univer. Texas Press, 1961. Pp. 397–412.

Downer, J. L. de C. Interhemispheric integration in the visual system. In V. B. Mountcastle (Ed.), *Hemispheric relations and cerebral dominance.* Baltimore: Johns Hopkins Press, 1962. Pp. 87–100.

Egger, M. D., & Flynn, J. P. Amygdaloid suppression of hypothalamically elicited attack behaviour. *Science,* 1962, **136**, 43–44.

Eliasson, S. Cerebral influence on gastric motility in the cat. *Acta physiol. Scand., Stockholm,* 1952, **26**(Suppl. No. 95), 1–70.

Elwers, M., & Critchlow, V. Precocious ovarian stimulation following hypothalamic and amygdaloid lesions in rats. *Amer. J. Physiol.,* 1960, **198**, 381–385.

Elwers, M., & Critchlow, V. Precocious ovarian stimulation following interruption of stria terminalis. *Amer. J. Physiol.,* 1961, **201**, 281–284.

Endröczi, E., & Lissák, K. Spontaneous goal-oriented motor activity related to the alimentary conditioned reflex behavior and its regulation by neural and hormonal factors. *Acta physiol. Hung.,* 1962, **21**, 265–283.

Feindel, W. Response patterns elicited from the amygdala and deep temporoinsular cortex. In D. E. Sheer (Ed.), *Electrical stimulation of the brain.* Austin: Univer. Texas Press, 1961. Pp. 519–533.

Feindel, W., & Gloor, P. Comparison of electrographic effects of stimulation of the amygdala and brain stem reticular formation in cats. *EEG clin. Neurophysiol.,* 1954, **6**, 389–402.

Feindel, W., & Penfield, W. Localization of discharge in temporal lobe automism. *AMA Arch. Neurol. Psychiat.,* 1954, **72**, 605–630.

Fernandez-deMolina, A., & Hunsperger, R. W. Affective reactions obtained by electrical stimulation of the amygdala. *J. Physiol.,* 1957, **138**, 29–30.

Fernandez-deMolina, A., & Hunsperger, R. W. Central representation of affective reactions in forebrain and brain stem: Electrical stimulation of amygdala, stria terminalis, and adjacent structures. *J. Physiol.,* 1959, **145**, 251–265.

Fernandez-deMolina, A., & Hunsperger, R. W. Organization of the subcortical system governing defense and flight reactions in the cat. *J. Physiol.,* 1962, **160**, 200–213.

Fonberg, E., The inhibitory role of amygdala stimulation. *Acta Biol. exp., Warsaw,* 1963, **23**, 171–180.

Fonberg, E., Brutkowski, S., & Mempel, E. Defensive conditioned reflexes and neurotic motor reactions following amygdalectomy

in dogs. *Acta Biol. exp., Warsaw,* 1962, 22, 51–57.

Fonberg, E., & Delgado, J. M. R. Avoidance and alimentary reactions during amygdala stimulation. *J. Neurophysiol.,* 1961, 24, 651–664.

Fox, C. A. Certain basal telencephalic centers in the cat. *J. comp. Neurol.,* 1940, 72, 1–62.

Fox, C. A. Amygdalo-thalamic connections in *Macaca mulatta. Anat. Rec.,* 1949, 103, 537–538. (Abstract)

Fuller, J. L., Rosvold, H. E., & Pribram, K. H. The effect on affective and cognitive behavior in the dog of lesions of the pyriform-amygdala-hippocampal complex. *J. comp. physiol. Psychol.,* 1957, 50, 89–96.

Gastaut, H. Corrélations entre le système nerveux végétatif et le système de la vie de relation dans le rhinencéphale. *J. Physiol. Pathol. gén.,* 1952, 44, 431–470.

Gastaut, H., Naquet, R., Vigouroux, R., & Corriol, J. Provocation de comportements emotionnels divers par stimulation rhinencéphalique chez le chat avec électrodes à demeure. *Rev. neurol., Paris,* 1952, 86, 319.

Gastaut, H., Vigouroux, R., Corriol, J., & Badier, M. Effets de la stimulation électrique (par électrodes à demeure) du complexe amygdalien chez le chat non narcosé. *J. Physiol. Pathol. gén.,* 1951, 43, 740–746.

Gastaut, H., Vigouroux, R., & Naquet, R. Comportements posturaux et cinétiques provoqués par stimulation sous-cortical chez le chat non anesthésié: Leur relation avec le «réflexe d'orientation.» *J. Psychol. norm. pathol.* 1952, 45, 257–271.

Gault, F. P., & Leaton, R. N. Electrical activity of the olfactory system. *EEG clin. Neurophysiol.,* 1963, 15, 299–304.

Gloor, P. Amygdala. In J. Field (Ed.), *Handbook of physiology.* Vol. 2. *Neurophysiology.* Washington, D. C.: American Physiological Society, 1960. Pp. 1395–1420.

Gloor, P., & Feindel, W. Affective behaviour and temporal lobe. In M. Monnier (Ed.), *Physiologie und Pathophysiologie des vegetativen Nervensystems.* Stuttgart: Hippokrates—Verlag, 1963. Pp. 685–716.

Goddard, G. V. The effect of amygdaloid stimulation on acquisition of a CER. *Amer. Psychologist,* 1963, 18, 473. (Abstract)

Green, J. D., Clemente, C. D., & DeGroot, J. Rhinencephalic lesions and behaviour in cats. *J. comp. Neurol.,* 1957, 108, 505–545.

Greer, M. A., & Yamada, T. Effect of bilateral ablation of the amygdala on endocrine function in the rat. In, Atlantic City Program, 41st meeting, Endocrine Society, 1959. P. 82. (Abstract)

Groethuysen, U. C., Robinson, D. B., Haylett, C. H., Estes, H. R., & Johnson, A. M. Depth electrographic recording of a seizure during a structured interview. *Psychosom. Med.,* 1957, 19, 353–362.

Heath, R. G., Monroe, R. R., & Mickle, W. A. Stimulation of the amygdaloid nucleus in a schizophrenic patient. *Amer. J. Psychiat.,* 1955, 111, 862–863.

Hilton, S. M., & Zbrozyna, A. W. Amygdaloid region for defense reactions and its efferent pathway to the brain stem. *J. Physiol.,* 1963, 165, 160–173.

Horvath, F. E. Effects of basolateral amygdalectomy on three types of avoidance behavior in cats. *J. comp physiol. Psychol.,* 1963, 56, 380–389.

Hunsperger, R. W. Affektreaktionen auf elektrische Reizung im Hirnstamm der Katze. *Helv. physiol. pharmacol. acta,* 1956, 14, 70–92.

Hunsperger, R. W. Les représentations centrales des réactions affectives dans le cerveau antérieur et dans le tronc cérébral. *Neuro-Chirurgie,* 1959, 5, 207–233.

Jasper, H. H., & Rasmussen, T. Studies of clinical and electrical responses to deep temporal stimulation in man with some considerations of functional anatomy. *Res. Publ. Ass. Nerv. Ment. Dis.,* 1956, 36, 316–334.

John, E. R., & Killam, K. F. Electrophysiological correlates of avoidance conditioning in the cat. *J. Pharmacol. exp. Ther.,* 1959, 125, 252–274.

John, E. R., & Killam, K. F. Electrophysiological correlates of differential approach-avoidance conditioning in cats. *J. nerv. ment. Dis.,* 1960, 131, 183–201.

Johnston, J. B. Further contributions to the study of the evolution of the forebrain. *J. comp. Neurol.,* 1923, 35, 337–481.

Kaada, B. R. Somato-motor, autonomic and electrocorticographic responses to electrical stimulation of "rhinencephalic" and other structures in primates, cat and dog: A study of responses from the limbic, subcallosal, orbito-insular, piriform and temporal cortex, hippocampus-fornix and amygdala. *Acta physiol. Scand., Stockholm,* 1951, 24 (Suppl. No. 83), 1–285.

Kaada, B. R., Andersen, P., & Jansen, J., Jr. Stimulation of the amygdaloid nuclear complex in unanesthetized cats. *Neurology,* 1954, 4, 48–64.

KAADA, B. R., JANSEN, J., JR., & ANDERSEN, P. Electrical stimulation of the hippocampus, cingulum bundle and related structures in unanesthetized cats. *Acta psychiat. neurol. Scand., Copenhagen,* 1952, **29**, 54.

KAADA, B. R., & URSIN, H. Further localization of behavioral responses elicited from the amygdala in unanesthetized cats. *Acta physiol. Scand., Stockholm,* 1957, **42**(Suppl. No. 145), 80–81. (a)

KAADA, B. R., & URSIN, H. Subcortical mechanisms in the "attention" response induced from the amygdala in cats. *Acta physiol. Scand., Stockholm,* 1957, **42**(Suppl. No. 145), 79–80. (b)

KING, F. A. Effect of septal and amygdaloid lesions on emotional behaviour and conditioned avoidance responses in the rat. *J. nerv. ment. Dis.,* 1958, **126**, 57–63.

KING, F. A., & MEYER, P. M. Effects of amygdaloid lesions upon septal hyperemotionality in the rat. *Science,* 1958, **128**, 655–656.

KING, H. E. Psychological effects of excitation in the limbic system. In D. E. Sheer (Ed.), *Electrical stimulation of the brain.* Austin: Univer. Texas Press, 1961. Pp. 477–486.

KLING, A., & HUTT, P. J. Effect of hypothalamic lesions on the amygdala syndrome in the cat. *Arch. Neurol. Psychiat.,* 1958, **79**, 511–517.

KLING, A., ORBACH, J., SCHWARTZ, N. B., & TOWNE, J. C. Injury to the limbic system and associated structures in cats. *Arch. gen. Psychiat.,* 1960, **3**, 391–420.

KLING, A., & SCHWARTZ, NEENA B. Effects of amygdalectomy on sexual behaviour and reproductive capacity in the male rat. *Federat. Proc.,* 1961, **20**, 335. (Abstract) (a)

KLING, A., & SCHWARTZ, NEENA B. Effects of amygdalectomy on feeding in infant and adult animals. *Federat. Proc.,* 1961, **20**, 335. (Abstract) (b)

KLINGLER, J., & GLOOR, P. The connections of the amygdala and of the anterior temporal cortex in the human brain. *J. comp. Neurol.,* 1960, **115**, 333–369.

KLÜVER, H. "The temporal lobe syndrome" produced by ablations. In G. E. W. Wolstenholme & C. M. O'Conner (Eds.), *Neurological basis of behavior.* Boston, Mass.: Little, Brown, 1958. Pp. 175–186.

KLÜVER, H., & BUCY, P. C. An analysis of certain effects of bilateral temporal lobectomy in the rhesus monkey, with special reference to "psychic blindness." *J. Psychol.,* 1938, **5**, 33–54.

KNIGGE, K. M. Adrenocortical response to immobilization in rats with lesions in hippocampus and amygdala. *Federat. Proc.,* 1961, **20**, 185. (Abstract)

KOIKEGAMI, H., DODO, T., MOCHIDA, Y., & TAKAHASHI, H. Stimulation experiments on the amygdaloid nuclear complex and related structures: Effects upon the renal volume, urinary secretion, movements of the urinary bladder, blood pressure and respiratory movements. *Folia psychiat. neurol. Jap.,* 1957, **11**, 157–207.

KOIKEGAMI, H., & FUSE, S. Studies on the functions and fiber connections of the amygdaloid nuclei and periamygdaloid cortex: Part 2. Experiment on the respiratory movements. *Folia psychiat. neurol. Jap.,* 1952, **5**, 188–197. (a)

KOIKEGAMI, H., & FUSE, S. Studies on the functions and fiber connections of the amygdaloid nuclei and periamygdaloid cortex: Part 2. Experiment on the respiratory movements. *Folia psychiat. neurol. Jap.,* 1952, **6**, 94–103. (b)

KOIKEGAMI, H., FUSE, S., HIROKI, S., KAZAMI, I., & KAGEYAMA, Y. On the inhibitory effect upon the growth of infant animals or on the obesity in adult cat induced by bilateral destruction of the amygdaloid nuclear region. *Folia psychiat. neurol. Jap.,* 1958, **12**, 207–223.

KOIKEGAMI, H., FUSE, S., YOKOYAMA, T., WATANABE, T., & WATANABE, H. Contributions to the comparative anatomy of the amygdaloid nuclei of mammals with some experiments of their destruction or stimulation. *Folia psychiat. neurol. Jap.,* 1955, **8**, 336–370.

KOIKEGAMI, H., KIMOTO, A., & KIDO, C. Studies on the amygdaloid nuclei and periamygdaloid cortex: Experiments on the influence of their stimulation upon motility of small intestine and blood pressure. *Folia psychiat. neurol. Jap.,* 1953, **7**, 86–108.

KOIKEGAMI, H., KUSHIRO, H., & KIMOTO, A. Studies on the functions and fiber connections of the amygdaloid nuclei and periamygdaloid cortex: Experiments on gastrointestinal motility and body temperature in cats. *Folia psychiat. neurol. Jap.,* 1953, **6**, 76–93.

KOIKEGAMI, H., YAMADA, T., & USUI, K. Stimulation of amygdaloid nuclei and periamygdaloid cortex with special reference to its effects on uterine movements and ovulation. *Folia psychiat. neurol. Jap.,* 1953, **8**, 7–31.

LESSE, H. Amygdaloid electrical activity during a conditioned response. *Proc. Int. Congr. Neurol. Sci., Brussels,* 1957, **1**. (a)

LESSE, H. Electrographic recordings of amygdaloid activity during a conditioned response. *Federat. Proc.*, 1957, **16**, 79. (Abstract) (b)

LESSE, H. Rhinencephalic electrophysiological activity during "emotional behavior" in cats. *Psychiat. res. Rep.*, 1960, **12**, 224–237.

LISSÁK, K., & ENDRÖCZI, E. Neurohumoral factors in the control of animal behavior. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning.* Toronto: Ryerson Press, 1961. Pp. 293–308.

LUNDBERG, P. O. Extrahypothalamic regions of the central nervous system and gonadotrophin secretion. *Proc. Int. Union Physiol. Sci., XXII Int. Congr.*, 1962, **1**, 615–619.

MACHNE, X., & SEGUNDO, J. P. Unitary responses to afferent volleys in amygdaloid complex. *J. Neurophysiol.*, 1956, **19**, 232–240.

MACLEAN, P. D., & DELGADO, J. M. R. Electrical and chemical stimulation of frontotemporal portion of limbic system in the waking animal. *EEG clin. Neurophysiol.*, 1953, **5**, 91–100.

MACLEAN, P. D., HORWITZ, N. H., & ROBINSON, F. Olfactory-like responses in pyriform area to non-olfactory stimulation. *Yale J. Biol. Med.*, 1952, **25**, 159–172.

MAGNUS, O., & LAMMERS, H. J. The amygdaloid-nuclear complex: Part 1. Electrical stimulation of the amygdala and periamygdaloid cortex in the waking cat. *Folia psychiat. neurol. neurochir. Neerl.*, 1956, **59**, 555–581.

MAHUT, H., & CORDEAU, J. P. Spatial reversal deficit in monkeys with amygdalohippocampal ablations. *Exp. Neurol.*, 1963, **7**, 426–434.

MANDELL, A. J., CHAPMAN, L. F., RAND, R. W., & WALTER, R. D. Plasma corticosteroids: Changes in concentration after stimulation of hippocampus and amygdala. *Science,* 1963, **139**, 1212.

MARTIN, J., ENDRÖCZI, E., & BATA, G. Effect of the removal of amygdalic nuclei on the secretion of adrenal cortical hormones. *Acta physiol. Hung.*, 1958, **14**, 131–134.

MASON, J. W. The central nervous system regulation of ACTH secretion. In H. H. Jasper et al. (Eds.), *Reticular formation of the brain.* Boston, Mass.: Little, Brown, 1958. Pp. 645–670.

MASON, J. W. Plasma 17-hydroxycorticosteroid levels during electrical stimulation in the amygdaloid complex in conscious monkeys. *Amer. J. Physiol.*, 1959, **196**, 44–48.

MASON, J. W., NAUTA, W. J. H., BRADY, J. V., & ROBINSON, J. A. Limbic system influences on the pituitary-adrenal cortical system. In, *Atlantic City Program, 41st meeting, Endocrine Society,* 1959. P. 29. (Abstract)

MASSERMAN, J. H., LEVITT, M., MCAVOY, T., KLING, A., & PECHTEL, C. The amygdalae and behaviour. *Amer. J. Psychiat.*, 1958, **115**, 14–17.

MEYER, D. R. Some psychological determinants of sparing and loss following damage to the brain. In H. E. Harlow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior.* Madison: Univer. Wisconsin Press, 1958. Pp. 173–192.

MISHKIN, M. Visual discrimination performance following partial ablations of the temporal lobe: II. Ventral surface vs. hippocampus. *J. comp. physiol. Psychol.*, 1954, **47**, 187–193.

MISHKIN, M., & PRIBRAM, K. H. Visual discrimination performance following partial ablations of the temporal lobe: I. Ventral vs. lateral. *J. comp. physiol. Psychol.*, 1954, **47**, 14–20.

MORGANE, P. J., & KOSMAN, A. J. Alterations in feline behaviour following bilateral amygdalectomy. *Nature, London,* 1957, **180**, 598–600.

MORGANE, P. J., & KOSMAN, A. J. A rhinencephalic feeding center in the cat. *Amer. J. Physiol.*, 1959, **197**, 158–162.

MORGANE, P. J., & KOSMAN, A. J. Relationship of the middle hypothalamus to amygdalar hyperphagia. *Amer. J. Physiol.*, 1960, **198**, 1315–1318.

MORIN, G., GASTAUT, H., VIGOUROUX, R., & ROGER, A. Comportement émotionnel et lésions expérimentales du rhinencéphale chez le chat. *CR Soc. Biol.*, 1952, **146**, 1959–1961.

MORRELL, F., ROBERTS, L., & JASPER, H. H. Effect of focal epileptogenic lesions and their ablation upon conditioned electrical responses of the brain in the monkey. *EEG clin. Neurophysiol.*, 1956, **8**, 217–234.

MORRIS, A. A. Temporal lobectomy with removal of uncus, hippocampus, and amygdala. *AMA Arch. Neurol. Psychiat.*, 1956, **76**, 479–496.

MULLAN, S., & PENFIELD, W. Illusions of comparative interpretation and emotion: Production by epileptic discharge and by electrical stimulation in the temporal cortex. *AMA Arch. Neurol. Psychiat.*, 1959, **81**, 269–284.

NAQUET, R. Sur les fonctions du rhinencéphale d'après les résultats de sa stimula-

tion chez le chat. Unpublished doctoral dissertation, Université d'Aix-Marseille, 1953.

Naquet, R. Effects of stimulation of the rhinencephalon in the waking cats. *EEG clin. Neurophysiol.*, 1954, 6, 711–712.

Olds, J., & Milner, P. M. Positive reinforcement produced by electrical stimulation of the septal area and other regions of the rat brain. *J. comp. physiol. Psychol.*, 1954, 47, 419–427.

Orbach, J., Milner, B., & Rasmussen, T. Learning and retention in monkeys after amygdala-hippocampus resection. *Arch. Neurol.*, 1960, 3, 230–251.

Penfield, W. The role of the temporal cortex in certain psychical phenomena. *J. ment. Sci.*, 1955, 101, 451–465.

Poirier, L. Anatomical and experimental studies on the temporal pole of the macaque. *J. comp. Neurol.*, 1952, 96, 209–248.

Porter, R. W., Conrad, D. G., & Brady, J. V. Some neural and behavioral correlates of electrical self-stimulation of the limbic system. *J. exp. Anal. Behav.*, 1959, 2, 43–55.

Pribram, K. H. The intrinsic systems of the forebrain. In J. Field (Ed.), *Handbook of physiology*. Vol. 2. *Neurophysiology*. Washington, D. C.: American Physiological Society, 1960. Pp. 1323–1372.

Pribram, K. H., & Bagshaw, M. Further analysis of the temporal lobe syndrome utilizing frontotemporal ablations. *J. comp. Neurol.*, 1953, 99, 347–375.

Pribram, K. H., & Kruger, L. Functions of the "olfactory brain." *Ann. N. Y. Acad. Sci.*, 1954, 58, 109–138.

Pribram, K. H., Lennox, M. A., & Dunsmore, R. H. Some connections of the orbito-frontotemporal, limbic and hippocampal areas of *Macaca mulatta*. *J. Neurophysiol.*, 1950, 13, 127–135.

Pribram, K. H., Mishkin, M., Rosvold, H. E., & Kaplan, S. J. Effects on delayed-response performance of lesions of dorsolateral and ventromedial frontal cortex of baboons. *J. comp. physiol. Psychol.*, 1952, 45, 565–575.

Pribram, K. H., & Weiskrantz, L. A comparison of the effects of medial and lateral cerebral resections on conditioned avoidance behavior in monkeys. *J. comp. physiol. Psychol.*, 1957, 50, 74–80.

Robinson, B. W., & Mishkin, M. Alimentary responses evoked from extrahypothalamic structures in *Macaca mulatta*. *Federat. Proc.*, 1961, 20, 327. (Abstract)

Robinson, E. The effect of amygdalectomy on fear-motivated behavior in rats. *J. comp. physiol. Psychol.*, 1963, 56, 814–820.

Rosvold, H. E., Fuller, J. L., & Pribram, K. E. Ablation of pyriform, amygdala, hippocampal complex in genetically pure strain cocker spaniels. In J. F. Fulton (Ed.), *Frontal lobotomy and affective behavior: A neurophysiological analysis*. New York: Norton, 1951. Pp. 80–82.

Rosvold, H. E., Mirsky, A. F., & Pribram, K. H. Influence of amygdalectomy on social behavior in monkeys. *J. comp. physiol. Psychol.*, 1954, 47, 173–178.

Sano, T. Motor and other responses elicited by electrical stimulation of the cat's temporal lobe. *Folia psychiat. neurol. Jap.*, 1958, 12, 152–176.

Sawa, M., & Delgado, J. M. R. Amygdala unitary activity in the unrestrained cat. *EEG clin. Neurophysiol.*, 1963, 15, 637–650.

Sawa, M., Ueki, Y., Arita, M., & Harada, T. Preliminary report on the amygdaloidectomy on psychotic patients, with interpretation of oral-emotional manifestation in schizophrenics. *Folia psychiat. neurol. Jap.*, 1954, 7, 309–329.

Schreiner, L., & Kling, A. Behavioral changes following paleocortical injury in rodents, carnivores and primates. *Federat. Proc.*, 1953, 12, 128. (Abstract) (a)

Schreiner, L., & Kling, A. Behavioral changes following rhinencephalic injury in cats. *J. Neurophysiol.*, 1953, 16, 643–659. (b)

Schreiner, L., & Kling, A. Effects of castration on hypersexual behavior induced by rhinencephalic injury in cats. *AMA Arch. Neurol. Psychiat.*, 1954, 72, 180–186.

Schreiner, L., & Kling, A. Rhinencephalon and behavior. *Amer. J. Physiol.*, 1956, 184, 486–490.

Schwartzbaum, J. S. Changes in reinforcing properties of stimuli following ablation of the amygdaloid complex in monkeys. *J. comp. physiol. Psychol.*, 1960, 53, 388–395. (a)

Schwartzbaum, J. S. Response to changes in reinforcing conditions of bar pressing after ablation of the amygdaloid complex in monkeys. *Psychol. Rep.*, 1960, 6, 215–221. (b)

Schwartzbaum, J. S. Some characteristics of "amygdaloid hyperphagia" in monkeys. *Amer. J. Psychol.*, 1961, 74, 252–259.

Schwartzbaum, J. S., & Pribram, K. H. The effects of amygdalectomy in monkeys on transposition along a brightness continuum. *J. comp. physiol. Psychol.*, 1960, 53, 396–399.

Scoville, W. B., Dunsmore, R. H., Liberson, W. T., Henry, C. E., & Pepe, A. Ob-

servations on medial temporal lobotomy and uncotomy in the treatment of psychotic states. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1953, **31**, 347–369.

SEGUNDO, J. P., NAQUET, R., & ARANA, R. Subcortical connections from the temporal cortex of the monkey. *AMA Arch. Neurol. Psychiat.*, 1955, **73**, 515–524.

SEN, R. N., & ANAND, B. K. Effect of electrical stimulation of the limbic system of brain ("visceral brain") on gastric secretory activity and ulceration. *Indian J. med. Res.*, 1957, **45**, 515–521.

SHEALY, C. N., & PEELE, T. L. Studies on amygdaloid nucleus of the cat. *J. Neurophysiol.*, 1957, **20**, 125–139.

SHEER, D. E. Emotional facilitation in learning situations with subcortical stimulation. In D. E. Sheer (Ed.), *Electrical stimulation of the brain.* Austin: Univer. Texas Press, 1961. Pp. 431–464.

SMITH, W. K. Non-olfactory functions of the pyriform-amygdaloid-hippocampal complex. *Federat. Proc.*, 1950, **9**, 118. (Abstract)

SPIEGEL, E. A., MILLER, H. R., & OPPENHEIMER, M. J. Forebrain and rage reactions. *J. Neurophysiol.*, 1940, **3**, 538–548.

STAMM, J. S., & KNIGHT, M. Learning of visual tasks by monkeys with epileptogenic implants in temporal cortex. *J. comp. physiol. Psychol.*, 1963, **56**, 254–260.

SUMMERS, T. B., & KAELBER, W. W. Amygdalectomy: Effects in cats and a survey of its present status. *Amer. J. Physiol.*, 1962, **203**, 1117–1119.

TERZIAN, H., & ORE, G. D. Syndrome of Klüver and Bucy reproduced in man by bilateral removal of the temporal lobes. *Neurology,* 1955, **5**, 373–380.

THOMPSON, A. F., & WALKER, A. E. Behavioral alterations following lesions of the medial surface of the temporal lobe. *Folia psychiat. neurol. neurochir. Neerl.*, 1950, **53**, 444–452.

THOMPSON, A. F., & WALKER, A. E. Behavioral alterations following lesions of the medial surface of the temporal lobe. *Arch. Neurol. Psychiat.*, 1951, **65**, 251–252.

TURNER, E. A. Cerebral control of respiration. *Brain,* 1954, **77**, 448–486.

UEKI, S., & DOMINO, E. F. Some evidence for a mechanical receptor in olfactory function. *J. Neurophysiol.*, 1961, **24**, 12–25.

URSIN, H., & KAADA, B. R. Functional localization within the amygdaloid complex in the cat. *EEG clin. Neurophysiol.*, 1960, **12**, 1–20. (a)

URSIN, H., & KAADA, B. R. Subcortical structures mediating the attention response induced by amygdala stimulation. *Exp. Neurol.*, 1960, **2**, 109–122. (b)

WALKER, A. E., THOMPSON, A. F., & McQUEEN, J. D. Behavior and the temporal rhinencephalon in the monkey. *Johns Hopkins Bull.*, 1953, **93**, 65–93.

WEISKRANTZ, L. Behavioral changes associated with ablation of the amygdala. *Amer. Psychologist,* 1953, **8**, 452. (Abstract)

WEISKRANTZ, L. Behavioral changes with ablation of the amygdaloid complex in monkeys. *J. comp. physiol. Psychol.*, 1956, **49**, 381–391.

WEISKRANTZ, L., & WILSON, W. A., JR. The effects of reserpine on emotional behavior of normal and brain-operated monkeys. *Ann. N. Y. Acad. Sci.*, 1955, **61**, 36–55.

WEISKRANTZ, L., & WILSON, W. A., JR. The effect of ventral rhinencephalic lesions on avoidance thresholds in monkeys. *J. comp. physiol. Psychol.*, 1958, **51**, 167–171.

WENDT, R., & ALBE-FESSARD, D. Sensory responses of the amygdala with special reference to somatic afferent pathways. In, *Physiologie de l'hippocampe.* Paris: Centre National de la Recherche Scientifique, 1962. Pp. 171–200.

WHITLOCK, D. G., & NAUTA, W. J. H. Subcortical projections from the temporal neocortex in *Macaca mulatta. J. comp. Neurol.*, 1956, **106**, 183–212.

WOOD, C. D. Behavioral changes following discrete lesions of temporal lobe structures. *Neurology,* 1958, **8**, 215–220.

WOOD, C. D., SCHOTTELIUS, B., FROST, L. L., & BALDWIN, M. Localization within the amygdaloid complex of anesthetized animals. *Neurology,* 1958, **8**, 477–480.

WOOD, J. W. Loss of aggressiveness in wild rats following lesions in the rhinencephalon. *Int. Physiol. Congr. Abstr.*, 1956, **20**, 978–979.

WURTZ, R. H. Self-stimulation and escape behavior in response to stimulation of the rat amygdala. Unpublished doctoral dissertation, University of Michigan, 1962.

WURTZ, R. H., & OLDS, J. Chronic stimulation of amygdaloid complex. *Federat. Proc.*, 1961, **20**, 336. (Abstract)

YAMADA, T., & GREER, M. A. The effect of bilateral ablation of the amygdala on endocrine function in the rat. *Endocrinology,* 1960, **66**, 565–574.

ZBROZYNA, A. W. Defense reactions from the amygdala and the stria terminalis. *J. Physiol.*, 1960, **153**, 27–28.

(Received August 21, 1963)

Psychological Bulletin
1964, Vol. 62, No. 2, 110–121

THREE THEORETICAL VIEWS OF SLANT PERCEPTION

HOWARD R. FLOCK

Hunter College of the City University of New York

3 current theories of slant perception are described. It is shown that the Gestaltist and Helmholtzian theories of slant perception depend on the shape-slant invariant and concomitantly on the familiarity of the presented shapes composing the surface (Helmholtzian) or on the goodness of the presented shapes (Gestaltist). In contrast the Gibsonian theory of slant perception depends on complex angular relations (optical texture gradients) in the optical texture (array of patterned light to the eye) without regard to the familiarity or goodness of the textural shapes composing the distal surface.

Before Gibson in the late 1940s started to emphasize the importance of the slant of a ground plane to the entire problem of space perception, the slant of a surface was a topic hardly referred to in the psychological literature. Indeed, to this day there is not to my knowledge a single review of the topic in the literature.

Whereas Gibson's theory of space perception has much to say about slant perception, two older theories of space perception, those of the Gestaltists and the Helmholtzians, have had little to say on the subject. Nevertheless, while both of the older theories have tended to ignore the problem of slant perception, they have been sufficiently explicit on the subject to make it known that their theories did comprehend the topic. In fact, the two theories implicitly assert that the topic of slant perception raises no special problems.

It is the purpose of this paper not so much to review the literature on slant perception, which is scanty, but to make explicit what the three major theories of space perception have to say on the subject of slant perception. At the same time I will specify, if not always discuss, theoretical and empirical problems indigenous to the specific theories.

Gestalt Theory of Slant Perception

In 1935 Koffka, following on the earlier work of Thouless (1931a, 1931b, 1932), Eissler (1933), and Klimpfinger (1933), related the problem of explaining slant perception to the perception of shape. This relationship came to be known as Koffka's *shape-slant invariant*.

Koffka (1935) took the position that a certain combination of shape and slant is invariant for a given retinal shape (p. 233). For a particular retinal projection (or retinal shape), seeing the shape in a certain way requires one to see the slant in a certain way. For the same retinal projection any change in apparent shape necessarily produces a corresponding change in apparent slant. Without this assumption, argued Koffka (1935), no intelligible theory of perception is possible (p. 230).

Koffka (1935) based the law of the invariant on two axioms. In the first (Axiom I), he said that supraliminally unequal retinal shapes cannot produce entirely equal perceptual effects (p. 230). Consider, for instance, a square that is slanted back into depth. It projects a retinal shape so that the vertical is shorter than the horizontal. If, despite these unequal, presumptively liminal,

retinal lengths, the subject sees a square, then, according to Koffka, he must also see the figure as slanted back into depth (or see the sides at different depths, an alternative which I ignore in the discussion below).

In the converse of this rule (Axiom II) Koffka (1935) stated that if two equal retinal shapes give rise to two different perceived shapes, they at the same time produce the impression that these two shapes are differently oriented (p. 229). Consider, for instance, a rectangle whose height to width is truly in the proportion of 2:1. Slant this rectangle back into depth so that the projected (retinal) height to width is 1:1. If these two equal retinal extents are seen to be of different lengths (for instance, seen to be in the true ratio of 2:1), then according to Axiom II the height of the figure must be seen as slanted back into depth.

Koffka's two axioms do not say, for some apparent shape and a fixed retinal shape, just what apparent slant should occur. Nevertheless, Koffka tacitly acknowledged the form of the systematic relationship between phenomenal shape and slant in his discussion of Eissler's (1933) and Klimpfinger's (1933) data. Like Stavrianos (1945) and Beck and Gibson later (1955), Eissler and Klimpfinger sought to determine whether the accuracies of shape and slant judgments covary, given a fixed retinal shape. Koffka, in evaluating their data, accepted this requirement of covariation and implied that the covariant relation has the form, $XY = K$, where X is apparent shape, Y is apparent slant, and K is the retinal projection of the shape-at-a-slant. Some implications of this relationship are developed below.

Consider a simple wire form, the figure, L, which is constituted of a horizontal shaft 1 unit long, a vertical shaft 2 units high, and an included angle of 90 degrees. If one defines the shape of an object as constant internal relations of position and proportionate distance (Oxford English Dictionary, 1914, p. 90), then the shape of the wire figure described above is given in part by the ratio of its height to its width (H/W).

In Figure 1 the L form is shown with an angle of slant, θ, so that its height (H) projects a frontal extent (t) of 1 unit, the same projected length as that for the nonslanted horizontal dimension (W). The angular sizes of H and W are given by angles α and β, respectively, and the corresponding retinal images by h and w. Finally, D specifies the physical distance between the eye and the form; and d specifies the dis-

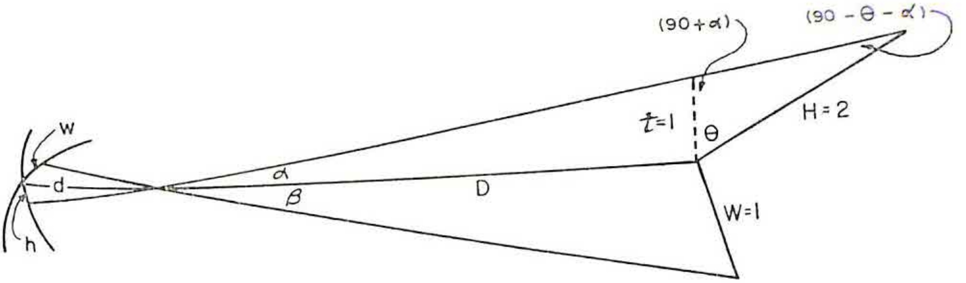


FIG. 1. Geometric basis for a shape-slant invariant. (Projection to a point of the angular sizes, α and β, and to the retina of the retinal extents, h and w, of an L form with height (H) = 2, width (W) = 1, included angle = 90 degrees, that is slanted by the angle, θ, into depth, and which is at a distance, D; d = distance between nodal point and retina; t = 1, and is the frontal projection of H.)

tance between the eye's nodal point and the retina. These constructions and some further included angles are given in Figure 1.

The mathematical form, $XY = K$, of Koffka's shape-slant invariant may now be developed as follows: By the law of sines, $H/\sin\ (90 + \alpha) = t/\sin\ (90 - \alpha - \theta)$; cross multiply, simplify, and solve for t; then $t = H\ [\cos\theta\ (1 - \tan\alpha \tan\theta)]$. Since in Figure 1, $h/d = t/D$ and $w/d = W/D$, then, substituting for D, $t = W\ (h/w)$. Substituting in the equation above for t and dividing through by W gives:

$$(H/W)\ [\cos\theta\ (1 - \tan\alpha \tan\theta)] = h/w \quad [1]$$

Equation 1 is in the form $XY = K$, where X is H/W, the apparent shape; Y is the expression, $\cos\theta\ (1 - \tan\alpha \tan\theta)$, the apparent slant; and K is h/w, the retinal shape. This equation provides a solution for apparent slant when values are given to the terms corresponding to the retinal shape and to the apparent shape. Moreover, the expression, $\tan\alpha \tan\theta$, approaches zero as the distance between the eye and the object increases. When D is 10 times longer than the size of the slanted dimension of the object, for example, disregarding the expression, $\tan\alpha \tan\theta$, introduces an error of approximately no more than 5% in the solution of θ.

That Equation 1 embodies Koffka's idea of the shape-slant invariant may be demonstrated by substituting Gestaltist terms for the terms of the equation. The term, H/W, would be replaced by the phrase, autochthonously generated phenomenal shape; the term, $\cos\theta\ (1 - \tan\alpha \tan\theta)$, by autochthonously generated phenomenal slant; and h/w by retinal process. Call this form of Equation 1, Equation 1a.

Equation 1a makes possible a direct empirical test of the truth of Koffka's hypothesis that the accuracies of phenomenal shape and slant judgments covary. For a given retinal process, so that the projected retinal height to width is 1.0, as above, if a physical shape of 2:1 is inaccurately reported as being 1.7:1, for example, then reported slant should correspondingly be in error.

Empirical tests of the shape-slant hypothesis (and its consequences) have yielded wide discrepancies between the accuracy of shape and slant judgments and even those friendly to the hypothesis have failed to demonstrate the results predicted by Koffka. (See Beck & Gibson, 1955; Eissler, 1933; Flock, 1964c; Klimpfinger, 1933; Langdon, 1953, 1955; Moore, 1938; Stavrianos, 1945. For an excellent review of this literature, see Epstein & Park, 1963.)

The relations described in Equation 1a raise a more crucial question for Gestalt theory. When a shape whose true proportions are 2:1 is slanted back so as to project a retinal ratio of 1:1, the retinal result is described by Gestalists as being symmetrical, regular, and dynamically stable. And yet if the perceiver is to see the physical shape correctly, the symmetrical retinal ratio must be transformed into a cortical ratio of 2:1. Since all perceptual theorists must finally account for the fact that perceptions generally are more or less veridical, Gestalt theory should give an account of how a good retinal process of 1:1 is transformed into a less good cortical process of 2:1. If a mechanism to account for such transformations of shapes does exist in Gestalt theory, then for a given retinal image and following upon a correct shape judgment, a correct slant judgment would necessarily follow from the solution of Equation 1a.

The only mechanism in Gestalt theory to account for shape perceptions is

that of the principle of *Prägnanz*. But it is not clear how that principle can be employed to explain a transformation from a stable retinal shape, for example, a shape with 1:1 proportions, to a less stable cortical shape, for example, a shape with 2:1 proportions.

There is, in addition, a difficulty between the hypothesized action of the principle of *Prägnanz* and the two axioms on which Koffka based the principle of a shape-slant invariant, as is explained below.

Prägnanz and Koffka's Axioms

When the Gestalist demonstrated the operation of the law of *Prägnanz,* he observed, for example, that certain supraliminal retinal proportions are cortically and therefore phenomenally transformed into some other set of proportions. But Axioms I and II require that concurrent with these changes some other effect must occur, for example, a perceptual change in the slant of an object or some part of an object. When an irregular shape is presented frontally, therefore, if some dimension is phenomenally more regularized than some other dimension (law of *Prägnanz*), then there must also be a phenomenal report of the irregular shape lying at a slant (Axioms I and II). Since the Gestalists were known to be careful phenomenological observers and yet failed to report such slant phenomena, one must conclude that the slant phenomena do not accompany the regularization process as is predicted by Axioms I and II. There is then an empirical contradiction between the hypothesized action of the principle of *Prägnanz* and the axioms that underlie the hypothesis of an invariant. This contradiction would seem to invalidate the truth value of either Axioms I and II or the law of *Prägnanz*. Further, if Axioms I and II are false, then the theoretical basis for the necessity of the shape-slant invariant disappears.

Koffka (1935) himself rejected the principle of regularization (one aspect of the changes observed from the application of the law of *Prägnanz*) as inapplicable to the problem of how a good retinal shape corresponding to some slanted figure can be transformed into a less good phenomenal shape (pp. 230–231). Nevertheless, this willingness to reject the application of *Prägnanz* to this problem does not erase the intrinsic contradictions in the theory.

Mechanisms for Generating Phenomenal Slant in Gestalt Theory

In the absence of a mechanism for generating correct phenomenal shapes, is there a mechanism in Gestalt theory that can account for correct phenomenal slants and, following upon Equation 1a, correct phenomenal shapes? Koffka proposed that certain features of the retinal distribution give rise to certain cortical processes which are characterized phenomenally as the major verticals and horizontals (framework) of the spatial perception. Frontal parallel planes are assumed to constitute some part of that framework.

Koffka said that a plane surface not in the frontal parallel plane of the framework sets up stresses in the brain field. But this assertion failed to say what there is about a physical plane at a slant that would generate stresses in the brain field. Koffka (1935) went on to add that the retinal shape projected by the figure in the slanted plane produces new forces that combine with the forces (stresses) of orientation (p. 231).

Koffka introduced two sets of new forces: the stresses of the nonnormally oriented plane and the new forces set up by the figure in the slanted plane. He did not say what there is in the proximal stimulus, produced by a shape

in a slanted plane, that would arouse new forces and stresses, and thereby appropriate perceptions. Insofar as I can follow the development of the Gestaltist's position, a slanted figure and/or a slanted plane project only some kind of retinal shape. That retinal shape occurs in relation with other retinal shapes. Some of these retinal shapes arouse neural processes that are isomorphic with the phenomenal reports of a spatial framework (verticals and horizontals). But this does not explain why other retinal shapes are seen as phenomenal figures at a slant. It is not clear, therefore, that these hypotheses about new forces give a satisfactory account of why slanted surfaces are seen as slanted.

Koffka maintained that the frontal position of a plane is a dynamically stable position. In assuming this he may have introduced further contradictions. Consider a perfectly regular checkerboard grid that is presented frontally. The projections of the squares to the retina take the form of a family of dynamically unstable angular trapezoids, not a family of dynamically stable angular squares as is implied by the Gestaltists. As in the case of the same surface when slanted, one should expect the retinal trapezoids to be cortically transformed into phenomenal squares and at the same time to yield the phenomenon of a slanted surface (principle of *Prägnanz* and Axioms I and II). It would seem that the normal state of a plane should be that simple state in which all of the good figures composing the surface project retinal images that are simple and stable. If checkerboard squares are distributed over a spherical surface with the eye at the center of the sphere, only then are the retinal shapes simple and stable angular squares.

This search of Gestalt theory for some mechanism that would predict the degree of accuracy of either apparent shape or apparent slant when a shape is at a slant has yielded little of substance. Koffka's explanatory concepts like stresses (corresponding to slanted planes) and new forces (corresponding to figures in slanted planes) were shown to have no specifiable physical correlates in the stimulus. Equally untenable was the basis for the shape-slant invariant on which the whole apparatus of explanation depended. It was shown that the idea of the invariant and its pair of axioms impose the necessity of demonstrating slant phenomena with occasions of the operation of *Prägnanz*. Since the Gestaltists have empirically failed to observe such slant phenomena accompanying instances of the action of *Prägnanz*, their own failures refute the principle of the shape-slant invariant. Moreover, empirical tests of the shape-slant invariant have failed to provide evidence for its necessity.

Disregarding many of these difficulties, one might insist that the Gestaltists do have a mechanism in the law of *Prägnanz* to hypothesize perceptions of shape. It might be argued that when objective shapes are regular and simple, regardless of their slant, perceptions of shape are more accurate than when objective shapes are irregular. Then, entering the degree of accuracy of the shape judgments into Equation 1a for the shape-slant invariant, one could predict the relative accuracy of the slant judgments. While this argument ignores difficulties, it at least satisfies the Gestaltist's need for a theory of slant perception and it allows for empirical predictions indigenous to that theory.

Perception of Surface Slant in Helmholtzian Theory

One of the most vigorous modern solutions to the various problems of explaining space perception dates back

to Helmholtz' *Physiological Optics* in 1866. Helmholtz, like most of his followers, gave no specific account of the perception of slant. In fact, the entire problem of explaining space perception was made subordinate to questions about the perception of distance and size. Nevertheless, Helmholtz' account of the perception of distance and size was sufficiently explicit to construct what he presumably might have advanced as his theory of slant perception. The basis for such a construction lies in his theory of distance perception and in his concept of previous acquaintance.

Helmholtz as well as his followers wrote as if a theory of space perception must start with an account of how judgments of distance and size are made in terms of metric units like meters or feet. Without regard to the correctness of that belief, Helmholtzian theories must be understood on the basis of that conviction. Since binocular convergence and parallax are relatively ineffective beyond 100 feet, the problem of explaining binocular perceptions of sizes and distances of objects was for Helmholtz the same as the problem of explaining monocular perceptions of size and distance.

The Concept of Previous Acquaintance

Helmholtz (1925) solved the problem of explaining monocular space perceptions, which for him were judgments, by making accurate perceptions depend on experience and on some previous acquaintance with the special nature of the perceived object (p. 282). On the basis of past experience with a particular object, he argued, the perceiver unconsciously makes an identification of the object from its retinal projection and attributes to the object some metrical size. These unconscious inferences together with the sensory inputs constitute the perception of the object and make possible the perception of its distance in a unit like feet.

The mathematical form in which this claim may be expressed is the essence of simplicity. Consider again Figure 1. If the L shape is presented frontally instead of at a slant, then the triangles formed by the terms D and H and d and h are similar, yielding Equation 2:

$$h/d = H/D \qquad [2]$$

One can replace the terms in Equation 2 with expressions indigenous to Helmholtz. The term H can be replaced by the term, attributed height of the object; the term D by judged or apparent distance; h by retinal extent; and d by distance between nodal point and retina. Call this form of Equation 2, Equation 2a.

The form of this equation is perfectly explicit in Helmholtz' (1925) comments on monocular space perceptions. He asserted that in monocular vision with a stationary head one cannot discern distance (p. 158). Nevertheless, if one is to have monocular space perceptions, then the distance from the eye of every point seen in the line of sight has to be known (p. 282). In order to solve the problem of how the perceiver would have such knowledge, Helmholtz depended on two terms: the retinal image and familiarity with the object. According to the Helmholtzian view one does discern with some degree of accuracy the retinal image, by which he meant the relative dimensions of the lines and angles which are projected from an object to a frontal plane or to the plane of the retina (p. 158). With the retinal image known to some degree of accuracy, the perceiver can then determine the distance of the object by depending on his previous acquaintance with the particular properties of the perceived object (p. 282).

It should not be subsequently forgotten what Helmholtz meant when he

spoke of acquaintance with the special nature of objects. He meant conscious or unconscious knowledge of metric sizes and angles of objects. When a phrase like "previous acquaintance" appears without its full context, there is the ever present danger that an apologist may read into it something more general than what Helmholtz meant. It can be demonstrated, for example, that all surfaces at a particular slant share in common a projection to the eye of complex angular relationships which identify their specific slant (see Flock, 1964b). Such angular relationships are not what Helmholtz (1925) meant by the special nature of the perceived object. What he did mean he unequivocally made clear:

> Knowing the size of an object, a human being, for instance, we can estimate the distance from us by means of the visual angle subtended, [my Equation 2a], or, what amounts to the same thing, by means of the size of the image on the retina. Persons in a landscape are particularly good objects for this purpose, because they are easy to recognize by their movements, they do not vary much in size, and we are familiar with them. . . . Houses, trees, plants, etc., may be used for the same purpose, but they are less satisfactory, because, not being so regular in size, such objects are sometimes responsible for bad mistakes. . . . This relation between distance and size is something that can only be acquired by long experience, and so it is not surprising that children are not very proficient at it, and are apt to make big mistakes [p. 283].

That familiarity with the object is essential to Helmholtz' (1925) theory of monocular space perception is apparent from his view of the converse of the familiar situation. Since he thought that the stuff of landscapes, rocks, or glaciers, for example, is unfamiliar, he claimed that photographs of such objects produce nothing perceptual but an unintelligible medley of grey spots to the eye (p. 285).

This same dependence on past experience with the particular object as the unit of analysis is to be found in the accounts of all theorists who have followed Helmholtz into the twentieth century, for example, Ames (1951), Brunswik (1944, 1956), Hastorff (1950), Ittelson (1951a, 1951b, 1960), Kilpatrick (1951), Vernon (1937), Woodworth and Schlosberg (1954), just to mention a few.

As an example of how closely the modern Helmholtzian paraphrases Helmholtz, consider Woodworth and Schlosberg's (1954) dependence on the concept of previous acquaintance in explaining distance perception. They maintained that if we know the real size of a visible object we have a good indication of its distance (p. 462). In their discussion of Ames' experiment with the distorted room, Woodworth and Schlosberg first derived an equation equivalent to Equations 2 and 2a above, making judged distance depend on unconscious estimates of an object's size. They spoke of whole families of such equations being neatly solved for distance. In Ames' room, according to Woodworth and Schlosberg (1954), the perceiver need only notice the familiar rectangular patterns of the room, for which the estimates of size will be reasonable sizes of such familiar objects as determined by prior experience (p. 489).

Equally literal in their acceptance of Helmholtz' concept of previous acquaintance were the perceptionists at Princeton and Dartmouth in the early 1950s. Ittelson (1951a), for example, criticizing Gibson's (1950b) concept of texture gradient, made space perceptions depend on that galaxy of assumptions which an individual makes as to the nature and significance of the objective world about him (p. 201). In the same article Ittelson (1951a) concluded that such a mechanism of as-

sumptions is inescapable if we are to account for the functioning of size as a cue to distance (p. 198).

Helmholtz' Theory of Slant Perception

Presumably Helmholtz would have explained slant perception in the same way that he accounted for distance perception, again depending on the concept of previous acquaintance. Consider again Equation 1 and make the following substitutions: substitute the term "unconscious attributions as to the true height and width of the object" for the term, H/W; the term "judged slant" for $\cos \theta \ (1 - \tan \alpha \ \tan \theta)$; and the term "retinal shape" for h/w. Call this form of Equation 1, Equation 1b. One may now solve Equation 1b for judged slant.

This solution of the problem of slant perception in terms of the Helmholtzian concepts of unconscious inferences and familiarity with the object and its size dimensions indicates that the explanation of slant perception for the Helmholtzian offers no new problems.[1] The writings of the neo-Helmholtzians support this observation. Woodworth and Schlosberg (1954), for example, devoted considerable space to distance perception. But on the subject of slant perception their discussion was cursory, leaving the problem of explanation, seemingly, to judgments of distance and size, and to unconscious inferences.

Equation 1b has another interest. It has the same form as Equation 1a for Koffka's shape-slant invariant. Equation 1b, therefore, is a statement of Koffka's shape-slant invariant in Helmholtzian terms. As for the Gestaltist, the shape-slant invariant is the Helmholtzian means of explaining the perception of slant.

Gestalt theory presumably depends on the good form of the textural shapes of a surface to produce correct phenomenal perceptions of slant. Helmholtzian theory depends on the concept of familiarity with the object, employs an unconscious inference about the specific or relative sizes of the shape's proportions, and then, also via the shape-slant invariant, generates a judgment of slant.

A comparison of Equations 1a and 1b clearly isolates what distinguishes a Helmholtzian from a Gestaltist. Both depend on the circumscribed retinal shape. Both make phenomenal slant a derived phenomenon dependent on the accuracy of judgments or perceptions of shape, or on the accuracy of autochthonous shape processes. They differ only in the mechanism by which they arrive at their predictions of what phenomenal shapes will occur, one emphasizing experience, the other physiology.

There are further similarities. For example, Helmholtz (1925) often implies that what is familiar is what the Gestaltist would call a good Gestalt (p. 285). To the extent that the two terms are synonymous, the predictions of the Gestaltist and the Helmholtzian would converge and be congruent.

Gibson on Perception of Slant

In contrast to the Helmholtzian and Gestalt theories, both of which tended to solve the problem of slant perception by making it dependent on shape perception, Gibson (1947, 1948, 1950b) proposed a new theory of space perception in the years after 1947. Unlike his predecessors Gibson made the problem of slant central to his theory of

[1] At one place Helmholtz (1925, p. 285) implied that judged slant could be derived directly from the retinal image and inferences about the true internal angles of objects. This approach would burden the organism even further, since in addition to having knowledge about the sizes of objects it would also have to "know" about their true shapes.

space perception. Correct space perceptions are said to depend on conditions which always imply the accurate perception of a surface at a slant. In order to give an account of correct slant perceptions, Gibson introduced two new concepts: optical texture and optical texture gradients. With these concepts as his analytic units, Gibson altered the empirical questions traditional to the area of space perception in general and of slant perception in particular.

Optical Texture

Gibson (1950b, 1961) has noted that to every point in space are projected rays of light from the surrounding surfaces. He has called the bundle of light rays converging on one of these points the optic array. Because adjacent elements in a particular surface absorb and reflect light differently, the reflected rays of light have different intensities and the optic array possesses a texture called the optical texture. Gibson (1950b, 1957) has noted also that a particular class of substances reflects luminous patterns that are unique to that class and these luminous patterns are identifiable within the optical texture:

> A natural optic array (the light to a point in space occupied by an eye) has a very complex structure. It contains patterns within patterns within patterns. . . . There is, in fact, a hierarchy of structure, and this reflects the various levels of structure of the physical world (mountains, trees, pebbles, crystals, molecules) [1957, p. 3].

Optical Texture Gradients

The principle that each class of substances reflects light uniquely implies that the optical texture possesses repetitive, congruent luminous patterns corresponding to substances of the same class; and possesses different congruent luminous patterns corresponding to substances of different classes (see Flock, 1964b). The presumptive identification of a set of congruent luminous patterns in the optical texture makes possible the definition of what Gibson (1950b) has called the optical texture gradient.

Gibson defined the optical texture gradient (some order of change like 1, 2, 3, 4; 1, 2, 4, 8, etc.) as the rate of increase of the density of elements. By the term elements Gibson has meant the luminous transitions in the optic array, which he refers to as adjacent cycles of light intensity (Gibson, 1950a, 1961; Gibson, Purdy, & Lawrence, 1955, note p. 1). These adjacent cycles (or changes) of intensity correspond to the light reflected from adjacent particles of differing matter within, for example, a surface texture.

By defining the optical texture gradient as the changing density of luminous transitions, Gibson has created an unnecessary difficulty for his theory. The structure of matter certainly is angularly much finer than the resolving powers of the human eye. This means that there are as many discriminable luminous transitions along one optic meridian as along any other, regardless of the position of a surface. Under most conditions, therefore, there can be no gradient of luminous transitions in the optical texture. By defining the optical texture gradient as the changing density of congruent luminous patterns, however, Gibson could have avoided many of these difficulties (see Flock, 1964b). That Gibson (1950b) might really have meant congruent luminous patterns rather than luminous transitions as the unit of the optical texture gradients is indicated by the way he has talked about optical texture gradients. He has described them in terms of the light reflected by particular classes of elements, like houses, fence posts, telegraph poles, fields, hills, etc.

Optical Texture Gradients as Stimuli for Slant Perceptions

For mathematical reasons Gibson (1950a) has argued that it is reasonable to identify the optical texture gradient with the slant of a distal surface to the line of regard, and thereby to call the texture gradient the stimulus for phenomenal slant. According to the geometry of perspective, he has noted, a flat surface projects a distribution of optical texture (congruent luminous patterns) that gets denser as the surface recedes. Moreover, the optical gradient of density is proportional to the angle of slant (p. 373).

The Nature of Gibson's Empirical Inquiry

Helmholtz invariably initiated a question about space perception with an inquiry into the mechanisms for identifying an object and making correct estimates of its size or distance in terms of units like feet and inches. In contrast, Gibson has argued that if two perceivers correctly see, for example, the relational sizes of trees, houses, people, etc., then presumably the space seen by the two perceivers would look approximately the same without regard to the fact that they might disagree widely as to the identity and metrical sizes and distances of the perceived objects. As a result, Gibson has focused on the information necessary for making relational judgments. That information, he has found, implies neither the identifying nor the metrical subjectivism of Helmholtzian theory. He would dispute, therefore, any claim that making identifications and metrical estimates are necessary conditions for space perceptions.

Whereas former theories had made the perception of slant dependent on the perception of shape, and the perception of shape dependent on behavioral factors like previous acquaintance or on physiological factors like the principle of *Prägnanz,* conditions all internal to the organism, Gibson has analyzed the proximal stimulus for variables which could replace the subjectivist constructs previously used. In this reaction to subjectivism, Gibson (1948) has never suggested an empty organism. He has asked only that one should attribute no more machinery to the organism than is necessary. While Gibson has had little to say about the subjective attributes which would be necessary to register his optical stimuli, he would presumably agree that the registration of stimuli like optical texture gradients does imply some organismic achievements.

Empirical Testing of Gibson's Theory of Slant Perception

Since 1950 Gibson and his associates have both theoretically and empirically demonstrated with varying dependence on mathematical expression that optic variables specify distal spatial properties and that these optic variables elicit appropriate judgments of spatial properties.

Of the various experimental papers emanating from Gibson's laboratory, by far the largest number of papers on a single topic has been concerned with the relation of optic variables to the phenomenal slant and change of slant of a surface: Bergman and Gibson (1959), Flock (1962), Gibson (1950a), Gibson and Cornsweet (1952), Gibson and Gibson (1957), Gibson, Gibson, Smith, and Flock (1959), Gibson, Olum, and Rosenblatt (1955), Gibson, Purdy, and Lawrence (1955), Purdy (1958), and von Fieandt and Gibson (1959).

Typically, the experiments of Gibson and his associates in contrast to those of the Helmholtzian or the Gestaltist present the perceiver with a display of

unfamiliar and irregular objects or textural elements (see, e.g., Flock, 1964a; Gibson, 1950a; Gibson & Gibson, 1957). The experimenter then demonstrates that the subject responds appropriately on the basis of the information present in the light to the eye (Gibson & Flock, 1962, e.g., explain the mountain illusion as an appropriate perceptual response to optic variables). This information typically involves complex angular relations. Organismic registration of such angular relations presumably requires the ability to pick up ratios of angles without regard to conscious or unconscious knowledge of the stuff whose angles are being regarded. Since the complex angular relations geometrically specify, for example, the distal surface slants, the registration of these relations and the concomitant phenomenon of slant do not entail the physiological autochthonous process of the Gestaltist.

Most recently, Flock (1964b) has described trigonometric angular variables in the optical texture which theoretically specify the distal slant of a physical surface to the motionless monocular perceiver both when the surface is motionless and when moving. In empirical demonstrations of the effects of the stimulus, Flock (1962, 1963, 1964c; Flock & Moscatelli, 1964) has shown that the visual slant judgments do covary with the optic variables but do not necessarily depend on Helmholtz' concept of previous acquaintance nor on the principle of *Prägnanz* as it is interpreted above to apply to slant perceptions. Flock (1964c) has demonstrated also that monocular slant judgments can be accurate even though shape judgments coincide with the retinal shapes rather than with the very different objective shapes.

Presumably a test of credibility among the three theories of slant perception depends on demonstrating whether the degree of accuracy of slant perceptions is best predicted by goodness of shapes (Gestaltist), familiarity of shapes (Helmholtzian), or optic information without regard to the status of the shapes (Gibsonian).

REFERENCES

Ames, A. A. Visual perception and the rotating trapezoidal window. *Psychol. Monogr.*, 1951, **65**(7, Whole No. 324).

Beck, J., & Gibson, J. J. The relation of apparent shape to apparent slant in the perception of objects. *J. exp. Psychol.*, 1955, **50**, 125–133.

Bergman, R., & Gibson, J. J. The negative after-effect of the perception of a surface slanted in the third dimension. *Amer. J. Psychol.*, 1959, 72, 364–374.

Brunswik, E. Distal focusing of perception: Size constancy in a representative sample of situations. *Psychol. Monogr.*, 1944, **56**(1, Whole No. 254).

Brunswik, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

Eissler, K. Die Gestaltkonstanz der Sehdinge bei Variation der Objekte und ihrer Einwirkunsweise auf den Wahrnemenden. *Arch. ges. Psychol.*, 1933, **88**, 487–551.

Epstein, W., & Park, J. N. Shape constancy: Functional relationships and theoretical formulations. *Psychol. Bull.*, 1963, **60**, 265–288.

Flock, H. R. *The monocular perception of surface slant.* (Doctoral dissertation, Cornell University) Ann Arbor, Mich.: University Microfilms, 1962, No. 62-2514.

Flock, H. R. Selective registration of information in visual judgment of surface slant. *Percept. mot. Skills,* 1963, **17**, 537–538.

Flock, H. R. Irregular electrostatic surface-textures for use in the point-source shadow-projector. *Amer. J. Psychol.*, 1964, in press. (a)

Flock, H. R. A possible optical basis for monocular slant perceptions. *Psychol. Rev.*, 1964, in press. (b)

Flock, H. R. Some conditions sufficient for accurate monocular perceptions of moving surface slants. *J. exp. Psychol.*, 1964, **67**, 560–572. (c)

Flock, H. R., & Moscatelli, A. Variables of surface texture and visual space perceptions. *Percept. mot. Skills,* 1964, in press.

GIBSON, E. J., GIBSON, J. J., SMITH, O. W., & FLOCK, H. R. Motion parallax as a determinant of perceived depth. *J. exp. Psychol.*, 1959, **58**, 40–51.

GIBSON, J. J. (Ed.) Motion picture testing and research. *AAF Aviat. Psychol. Program res. Rep.*, 1947, No. 7.

GIBSON, J. J. Studying perceptual phenomena. In T. G. Andrews (Ed.), *Methods of psychology*. New York: Wiley, 1948. Ch. 6.

GIBSON, J. J. The perception of visual surfaces. *Amer. J. Psychol.*, 1950, **63**, 367–384. (a)

GIBSON, J. J. *The perception of the visual world*. Boston, Mass.: Houghton Mifflin, 1950. (b)

GIBSON, J. J. The information contained in light. Paper read at Symposium on Space Perception, International Congress of Psychology, Brussels, July 1957.

GIBSON, J. J. Ecological optics. *Vis. Res.*, 1961, **1**, 253–262.

GIBSON, J. J., & Cornsweet, J. The perceived slant of visual surfaces: Optical and geographical. *J. exp. Psychol.*, 1952, **44**, 11–15.

GIBSON, J. J., & FLOCK, H. R. The apparent distance of mountains. *Amer. J. Psychol.*, 1962, **75**, 501–503.

GIBSON, J. J., & GIBSON, E. J. Continuous perspective transformations and the perception of rigid motion. *J. exp. Psychol.*, 1957, **54**, 129–138.

GIBSON, J. J., OLUM, P., & ROSENBLATT, F. Parallax and perspective during aircraft landings. *Amer. J. Psychol.*, 1955, **68**, 372–385.

GIBSON, J. J., PURDY, J., & LAWRENCE, L. A method of controlling stimulation for the study of space perception: The optical tunnel. *J. exp. Psychol.*, 1955, **50**, 1–14.

HASTORFF, A. H. The influence of suggestion on the relationship between stimulus size and perceived distance. *J. Psychol.*, 1950, **29**, 195–217.

HELMHOLTZ, H. *Physiological optics*. Vol. 3. (3rd ed. trans. by J. P. Southall) New York: Optical Society of America, 1925.

ITTELSON, W. H. Size as a cue to distance: Radial motion. *Amer. J. Psychol.*, 1951, **64**, 188–202. (a)

ITTELSON, W. H. Size as a cue to distance: Static localization. *Amer. J. Psychol.*, 1951, **64**, 54–67. (b)

ITTELSON, W. H. *Visual space perception*. New York: Springer, 1960.

KILPATRICK, F. P., & ITTELSON, W. H. Three demonstrations involving the visual perception of movement. *J. exp. Psychol.*, 1951, **42**, 394–402.

KLIMPFINGER, S. Über den Einfluss von intentionaler Einstellung und Übung auf die Gestaltkonstanz. *Arch. ges. Psychol.*, 1933, **88**, 551–598.

KOFFKA, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.

LANGDON, J. Further studies in the perception of a changing shape. *Quart. J. exp. Psychol.*, 1953, **5**, 89–107.

LANGDON, J. The role of spatial stimuli in the perception of shape. Part I. *Quart. J. exp. Psychol.*, 1955, **7**, 19–27.

MOORE, W. E. Experiments on the constancy of shape. *Brit. J. Psychol.*, 1938, **29**, 104–116.

PURDY, W. C. *The hypothesis of psychophysical correspondence in space perception*. (Doctoral dissertation, Cornell University) Ann Arbor, Mich.: University Microfilms, 1958, No. 58-5594.

STAVRIANOS, B. K. The relation of shape perception to explicit judgments of inclination. *Arch. Psychol.*, 1945, No. 296, 1–94.

THOULESS, R. S. Phenomenal regression to the real object. Part I. *Brit. J. Psychol.*, 1931, **21**, 339–359. (a)

THOULESS, R. S. Phenomenal regression to the real object. Part II. *Brit. J. Psychol.*, 1931, **22**, 1–29. (b)

THOULESS, R. S. Individual differences in phenomenal regression. *Brit. J. Psychol.*, 1932, **22**, 216–241.

VERNON, M. D. The perception of distance. *Brit. J. Psychol.*, 1937, **28**, 1–11, 115–149.

VON FIEANDT, K., & GIBSON, J. J. The sensitivity of the eye to two kinds of continuous transformation of a shadow-pattern. *J. exp. Psychol.*, 1959, **57**, 344–347.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. (Rev. ed.) New York: Holt, 1954.

(Received September 10, 1963)

Psychological Bulletin
1964, Vol. 62, No. 2, 122–131

SYSTEMATIC EFFECT OF RANDOM ERROR IN THE YOKED CONTROL DESIGN [1]

RUSSELL M. CHURCH
Brown University

The yoked control design has been employed to determine whether the effect of an event on a response is related to the temporal relationship between the event and that response. A number of specific examples of this design are discussed, and the conclusion is reached that the results are necessarily ambiguous. If the event does have an effect on behavior, systematic differences between the experimental group and its matched control group will emerge: (a) if the temporal relationship between response and event is relevant or (b) if various sources of random error are present, such as individual differences in the effectiveness of the event.

In an instrumental learning situation an experimental subject (*S*) receives an event as a function of a response. For example, each lever press of a rat may result in a slight increase in illumination or a mild shock of short duration. This event may affect some measured behavior, such as the rate of lever pressing. For some events it is not certain whether the temporal relationship between the event and the response is relevant to the observed effect, that is, whether we are dealing with learning or with some less specific process such as activation or sensitization. One method of establishing the relevance of this temporal relationship involves the use of the yoked control design.

The yoked control design involves a number of *S*s that are paired on some basis. One of the two members of each pair is randomly selected as the experimental *S*; the other member is selected as the control *S*. The experimental *S* is put into an instrumental learning situation in which an event occurs if it makes a particular response; the control *S* is put into a situation in which an equivalent event occurs if its yoked experimental *S* makes the response. The critical characteristic of this design is that an experimental *S* determines the presentation of events to its control *S* so that the experimental and control *S*s receive the same number and temporal distribution of events (Figure 1). Apparently the only difference between them is that the experimental *S*, but not the control *S*, always receives the event immediately after a response.

Consider a situation in which the null hypothesis is true (i.e., the temporal relationship between the response and the event is irrelevant), but one in which the event does have an effect on behavior. Using the yoked control design in such a situation the unwary experimenter would expect to falsely reject the null hypothesis with a probability equal to the level he had set for the Type I error. Unfortunately, he may falsely reject the null hypothesis more often if there are individual differences among *S*s that are randomly assigned to the experimental and control groups. As in other experimental designs individual differences result in random errors, but in the yoked control design individual differences also

[1] Preparation of this article was facilitated by Research Grant MH-2903 from the National Institute of Mental Health, United States Public Health Service.

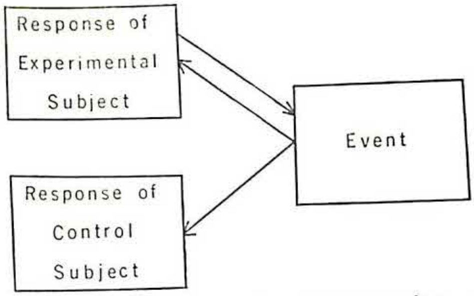


FIG. 1. Diagrammatic representation of the yoked control design. (Although both experimental and control *Ss* receive the same number and temporal distribution of events, only the experimental *S* produces events. Any difference in the behavior of the experimental and control *Ss* is assumed to be a result of this difference.)

may produce a constant error that cannot be reduced by increasing the number of observations. The following argument illustrates the point that a reliable difference in behavior between the experimental and control group may be due to the temporal relationship between the response and the event, or it may be due merely to individual differences.

INDIVIDUAL DIFFERENCES IN THE YOKED CONTROL DESIGN

Classical and Instrumental Eyelid Conditioning. The percentage of conditioned eyelid responses is higher under a classical procedure in which an unconditioned stimulus (UCS) is presented on each trial than under an instrumental avoidance procedure in which the UCS is presented only on trials that *S* fails to make a response (Kimble, Mann, & Dufort, 1955; Logan, 1951; Moore & Gormezano, 1961). Compared with *Ss* trained with the classical procedure, *Ss* trained with the instrumental avoidance procedure, (*a*) receive fewer presentations of the UCS, (*b*) receive the UCS on an irregular pattern of trials, and (*c*) avoid the UCS if they make a particular response. The yoked control design, by equating *Ss* with respect to the number and pattern of reinforcements, appears to be the ideal design to determine whether or not the correlation between response and reinforcement is related to the percentage of conditioned responses. On each trial an *S* in the instrumental avoidance group would receive a puff of air (UCS) $\frac{1}{2}$ second after conditioned-stimulus (CS) onset unless it made an eyelid response of sufficient magnitude and latency. The control *S* would receive a UCS $\frac{1}{2}$ second after CS onset if its matched experimental *S* received a UCS on that trial. Presumably the only difference between the two groups would be that the response and reinforcement were correlated in the instrumental avoidance group, but they are independent in the matched control group. Moore and Gormezano (1961) have performed this experiment, and they found that the instrumental avoidance group had a greater percentage of conditioned eyelid responses than a classical group equated for number and pattern of reinforcements. This result, as we will show, does not demonstrate the relevance of the temporal relationship between the eyelid response and the reinforcing event.

Let us suppose that classical conditioning occurs (the event does have an effect on behavior), but that the correlation between response and event is irrelevant. Thus, by assumption, the null hypothesis is true. Let us also make the eminently reasonable assumption that there are individual differences in the effectiveness of the event, for it is certainly well established that there are large differences in the percentage of conditioned responses made by various *Ss* under similar treatments. To simplify the exposition, let us assume that there are only two kinds of *Ss*, those for which the event is effective

in producing eyelid conditioning and those for which it is not (Table 1).

If an experimental *S* for which the event is effective is matched with a control *S* for which it is equally effective, no reliable difference in the percentage of conditioned responses will appear (Case 1). Similarly, no reliable difference will appear in those cases in which the event is ineffective for both *S*s (Case 4). Differences between the groups, however, appear in the second and third cases of Table 1. If an experimental *S* that is affected by the event is matched with a control *S* that is not, the experimental *S* will make substantially more responses than its matched control. If, on the other hand, an experimental *S* that is not affected by the event is matched with a control *S* that is affected, neither will make many responses. The experimental *S* will rarely make a response (by assumption) and the control *S* will not make many responses either because it only occasionally receives an event. Thus a significant difference in behavior between the experimental and control group may be due simply to individual differences in the ease of classical conditioning.

To simplify the exposition, each *S* was considered to be "affected" or "not affected" by the event. Of course, it is more reasonable to assume that the degree to which an *S* will be affected by an event is a continuous variable.

TABLE 1

INDIVIDUAL DIFFERENCES IN THE EFFECTIVENESS OF AN EVENT

Case	Experimental	Control
1	Effective	Effective
2	Effective	Ineffective
3	Ineffective	Effective
4	Ineffective	Ineffective

The systematic effects described above, however, also apply in this more general case. For experimental *S*s the frequency of events is related to their effectiveness; a similar relationship is not to be found in the case of control *S*s. Thus, as Moore and Gormezano (1961) state in the case of instrumental eyelid conditioning:

> The superior performance of *S*s in the instrumental procedure appears to be due then, to their receiving the UCS when *needed,* i.e., when the CR fails to occur, whereas, no such consistent relationship exists for the classical partial reinforcement procedure [p. 558].

Thus, the instrumental procedure may be a more effective method of training than the matched classical control, but it is more effective not because of the close temporal association between a response and a reinforcing event, but because instrumental training allocates events to *S*s that are most likely to profit from them.

Positive Reinforcers. Rats that receive a slight increase in illumination immediately following a lever response make significantly more such responses than rats that do not receive this event. The light may serve to reinforce the immediately preceding response, or it may serve to increase general activity, which could result in an increase in the measured rate of lever responding. To distinguish between these alternatives, Kling, Horowitz, and Delhagen (1956) used a yoked control design. Rats were paired on the basis of the frequency that they touched the lever and one *S* from each pair was randomly assigned to the experimental group. Whenever an *S* in the experimental group touched the lever, it received an increase in illumination and the *S* in the control group with which it was matched also received the same increase in illumination. On the first day of this

treatment experimental *S*s made significantly more responses than their matched control *S*s (Wilcoxon's signed ranks test, $p < .01$). Although these results appear to support the view that the light served as a reinforcing stimulus, they do not necessarily exclude the "activity" hypothesis. Using the same argument developed in the case of eyelid conditioning, we may conclude that if the event (light onset) increased the activity of some *S*s more than others, experimental *S*s might have responded significantly more often than the control *S*s with which they were matched.

Studies of secondary reinforcement are concerned with the reinforcing property of a previously neutral event (e.g., the sounding of a buzzer) after it has been paired with positive reinforcement. The *S*s that receive the event after each response may make more responses than *S*s that do not receive the event either because the event served as a reinforcer of the response or because it served as an arouser of activity. To distinguish between these alternatives a number of investigators have employed the yoked control design (Clayton & Savin, 1960; Crowder, Gay, Bright, & Lee, 1959; Crowder, Gay, Fleming, & Hurst, 1959; Crowder, Gill, Hodge, & Nash, 1959; Crowder, Morris, & McDaniel, 1959; Fox & King, 1961; Wyckoff, Sidowski, & Chambliss, 1958; Zimmerman, 1959). The results of studies in which experimental *S*s responded significantly more frequently than their matched control *S*s could be due to individual differences in the effectiveness of the event as an arouser of activity.

Negative Reinforcers. In many experiments it has been demonstrated that the presentation of a brief noxious stimulus, usually electric shock, immediately after the occurrence of a response results in suppression of the response. Does this response suppression occur because the punishing event occurred shortly after the response or does it occur because of the general suppression of behavior resulting from the presentation of the noxious stimuli? Although the yoked control procedure would serve to equalize the number and temporal distribution of punishing events, a reliable difference between the experimental and control groups could emerge even if the temporal relationship between response and punishment were irrelevant (Church, 1963). If, for example, the experimental *S* requires a few more punishments than its matched control *S* before it ceases to respond, there will be only a small difference in favor of the experimental *S*. If the control *S* requires a few more punishments than its matched experimental *S*, however, the control *S* will never cease responding.

Bindra and Anchel (1963) used a yoked control design to demonstrate that "immobility" could be used as an avoidance response. Shock was applied to the experimental animal whenever it was moving, and a matched control *S* received shock whenever the experimental animal received it. The assumption was made that if the rats remained immobile as a result of unconditioned "freezing" to the shock, the experimental *S*s and their matched control *S*s should be indistinguishable, but if immobility could be considered as a response reinforced by shock termination, then the experimental *S*s would remain immobile more than their matched control *S*s. Unfortunately, if *S*s developed a freezing pattern of response to shock at different rates, a significant difference could emerge between the groups even if the contingency between the response and shock termination were irrelevant. The familiar asymmetry also

appears in demonstrations of superior cardiac conditioning by *S*s in an instrumental avoidance group than by *S*s in a yoked classical control group (Shearn, 1962).

Do animals develop "fear from a sense of helplessness"? Mowrer and Viek (1948) compared the number of eating inhibitions of rats that could escape from an electric shock with those of rats that received equivalent durations of shock from which they could not escape. Experimental *S*s were presented with food on the end of a stick and, 10 seconds later, they received a shock from which they could escape by jumping into the air. Each experimental *S* was matched with a control *S* that received the presentation of food on the stick for 10 seconds and then shock of the same intensity and duration as that received by its matched experimental *S* on that trial. The only obvious difference in treatment was that the experimental *S* could escape from the shock by making a particular escape response while termination of the shock for the control *S* was not related to any particular response. Mowrer and Viek found that the number of eating inhibitions were significantly greater in the control group than in the experimental group, and they concluded, "A painful stimulus never seems so objectionable if one knows how to terminate it as it does if one has no control over it." Unfortunately, once again, individual differences in the effectiveness of the event could produce reliable differences between the experimental group and its matched control group. Let us only make the additional, but reasonable, assumption that the escape responses of *S*s for which the shock is particularly effective will be more rapid than those of *S*s for which the shock is relatively ineffective. Then, by referring back to Table 1, it will be noted that in Case 2 the experimental *S* will respond rapidly and thus the shock will be of brief duration to both it and its matched control *S*, but in Case 3 the experimental *S* will respond more slowly since the shock is relatively ineffective, and the control *S* for which the shock is effective will receive considerable noxious stimulation. This alone could produce a reliable difference between the experimental group and the matched control group. Although Gibson (1957) successfully replicated the Mowrer-Viek experiment, several studies have not found a difference in the amount of fear produced by escapable and nonescapable shocks (Brimer & Kamin, 1963; Stanley & Monkman, 1956), and one study (Brady, Porter, Conrad & Mason, 1958) found experimental monkeys that could avoid a shock by making a lever response developed more ulcers than their matched control *S*s. (This study, however, did not strictly employ a yoked control design since the *S*s were not randomly assigned to the two groups.) Presumably, in these instances, the extent of individual differences in the effectiveness of the event was not sufficiently great to produce reliable differences between the experimental group and its matched control group.

Sawrey and Weisz (1956) demonstrated that hungry and thirsty rats that receive electric shock whenever they eat or drink during repeated 47-hour sessions develop substantially more ulcers than unshocked controls. Are the ulcers a result of conflict, or merely of shock? Sawrey, Conger, and Turrell (1956) performed an experiment in which each experimental *S* that was in a conflict situation was paired with a control *S* that received equivalent shocks. The results were that the *S*s in conflict developed significantly

more ulcers than their matched controls. Let us assume that the event (shock to a deprived rat) is sufficient to produce ulcers, but that there are individual differences in susceptibility to ulceration. Systematic difference between the groups should emerge that are unrelated to conflict unless the unreasonable assumption is made that a rat suffering with ulcers is as likely to enter an electrified grid to obtain food and water as a normal rat.

General Statement Regarding Individual Differences. Although we have emphasized individual differences in the effectiveness of an event, other differences between the individuals can also produce the systematic effects described above. A reliable difference between the experimental group and its matched control group may arise from individual differences in the initial probability of response, in the effectiveness of the event, or in the final asymptotic level. Consider, for example, a simple linear operator of the form $p_{n+1} = p_n + \theta(\lambda - p_n)$. If there are individual differences in the initial probability of a response (p_o), in the effect of the event on the rate of change of the probability of a response (θ), or in the asymptotic level of performance (λ), systematic differences will emerge between an experimental group and its matched control group.

An Extension of the Yoked Control Design

In all of the examples we have considered so far, the control *S* received the event at the same instant in experimental time as its matched experimental *S*. More generally, the control *S* receives an event (*a*) if the experimental *S* received the event and (*b*) if it makes a specified response. The response required of the control *S* is usually more inclusive than that required of the experimental *S*, but it does not necessarily include all conceivable responses. For example, we will now consider situations in which the experimental *S* will receive an event if it makes a response with a certain speed requirement and the control *S* will receive an event (*a*) if the experimental *S* received the event and (*b*) if it makes the response at any speed. We may anticipate that the systematic influence of individual differences will remain in this extension of the yoked control design.

Positively Correlated Reinforcement. Ferster and Skinner (1957) used a yoked control procedure to demonstrate that the high response rates of pigeons on variable-ratio schedules of reinforcement are due to differential reinforcement of rate, rather than to increased frequency of reinforcement. Each *S* that was on a variable-ratio schedule was matched with a control *S* that received a reinforcement for its first response after its matched experimental *S* received a reinforcement. The experimental *S*s responded more rapidly than their matched control *S*s, despite the fact that they received the same number of reinforcements in each session. Although it is possible that differential reinforcement of response rate increased the response rate of the experimental *S*s, a reliable difference between an experimental group and its matched control group might equally well have appeared if the number of reinforcements that an experimental *S* received at a given session were a function of its response rate at the previous session. Response rate is positively related to the frequency of reinforcement and, presumably, there are individual differences in the magnitude of this relationship. In the yoked control design reinforcements are allocated to *S*s

in the experimental group that have demonstrated that they will profit from them; a similar allocation is not made in the case of the control group. Thus, if there are individual differences in the relationship between reinforcement frequency and response rate, systematic differences in response rate will emerge between experimental *S*s and their matched control *S*s. The same problem appeared in studies of competitive schedules of reinforcement (Church, 1961, 1962). Although the competitive allocation of reinforcements was assumed to be effective because of the differential reinforcement of response rate, it is now clear that there could have been a systematic difference between the experimental group and its yoked control group even if the differential reinforcement of rate was ineffective in changing the response rate.

Negatively Correlated Reinforcement. Logan (1960) has emphasized the importance of the correlation between running speed of rats in the straight alley and the conditions of reinforcement. Although experimental rats that were reinforced only if they reached the goal box before some time criterion did not learn to run significantly faster than their matched control *S*s, experimental rats that were reinforced only if they reached the goal box after some time criterion (e.g., 5 seconds) learned to run to the goal box significantly more slowly than their matched control *S*s (Logan, 1960, p. 165). The difference between the experimental group and its matched control group in the speed of running may be a result of the correlation between reinforcement and a particular running speed, or it may be a result of the method of allocation of the number of reinforcements to *S*s. This same ambiguity is present in the interpretation of the effects of competitive allocation of reinforcement to the slower of two rats in adjacent lever boxes (Church, 1961) or in adjacent runways (Carnathan & Church, 1964).

Variability within *S*s and the Yoked Control Design

If individual differences were the only source of random variation in the yoked control design there would be a simple way of making an unbiased estimate of the treatment effect. This would be to disregard all magnitudes of differences between the experimental *S* and its matched control *S*, and to consider only the direction of the difference between them. Under the null hypothesis, individual differences between *S*s that are randomly assigned to the experimental group and the matched control group would result in no systematic difference in the number of *S*s that are more affected in one group than the other. Unfortunately, it can be shown that even a sign test results in systematic bias if one assumes that there are within-*S* differences of a random nature. Even if the average effect of an event is the same for an experimental *S* and its matched control *S*, for a variety of reasons, one or the other *S* may be more affected during any given block of time. If this is the case, then all the statements made for a number of pairs of *S*s also apply to each single pair of *S*s over a number of trials. Consequently variability within *S*s may result in a systematic difference in the number of experimental *S*s that are more affected than their matched control *S*s.

Let us reconsider the comparison of instrumental avoidance eyelid condititioning with a classical matched control group, now with the assumption that, although there are no individual differences in the ease of classical conditioning, there is some variability from moment to moment in the effectiveness of a classical conditioning trial. If the trial is administered at an instant during which only the experimental *S* will

reach 100% criterion, the control *S* will have no further chance to become conditioned. If, on the other hand, the trial is administered at an instant during which only the control *S* will reach 100% criterion, the experimental *S* will receive further trials with the UCS. Thus each experimental *S* will continue to receive the events until it is conditioned, that is, it will receive whatever number of events it needs to become conditioned. On the other hand, some control *S*s will receive more events than necessary to become conditioned while others will not receive enough. Consequently, within-*S* variability among otherwise identical *S*s may produce a systematic difference between an experimental group and its matched control group. As a second example of the systematic effect of variability within *S*s, let us reconsider the experiments investigating "fear from a sense of helplessness" (Mowrer & Viek, 1948). Presumably there will be some variation from one moment to the next in the severity of the shock, as a result of variation in density and locus of the shock, in the degree of contact with the grid, etc. Whenever the experimental *S* receives a particularly painful shock, it may be able to escape quickly, but its matched control *S* is in a less fortunate position. Sometimes a particularly painful shock to the control *S* will be terminated by its matched experimental *S*, but this is not always the case. It may be that the "helpless" control *S* shows more eating inhibition due to the lack of correlation between a response and shock termination or it may be a result of moment to moment variation in the severity of the shock. These examples illustrate a point which may be applied to each of the other examples previously considered: A systematic difference between the experimental group and its matched control group may be produced by variability within *S*s, even in the absence of any individual differences.

One modification of the yoked control design involves a continual redefinition of the "response" so that the experimental *S* (and, therefore, its matched control *S*) receives the event close to some fixed percentage of trials. For example, Logan (1960) reinforced the running response of an experimental rat in a straight alley if it was faster than the median of its previous day's performance, and Church and Carnathan (1963) reinforced the lever-press response of an experimental rat to an auditory signal if the latency was less than the median of its previous day's performance. In these experiments each *S* in the experimental group and in the matched control group received reinforcement on approximately 50% of the trials. Here, then, there is no problem of individual differences, but systematic differences may emerge if there are within-*S* differences in the effectiveness of the event. In most situations it is unreasonable to assume that a given *S* is equally affected by an event in each block of time because of such factors as inattention and fear. If *S*s are more affected by the events in blocks of time when they are responding rapidly, events would be delivered to experimental *S*s when they would be most effective. Since there is usually no quantitative estimate available of the extent of the variations in effectiveness of the event, there does not seem to be any statistical analysis possible to separate the effects of differential reinforcement from sources of random error in the yoked control design.

Alternatives to the Yoked Control Design

Alternative experimental designs are available to test the research hypotheses that have been investigated with the

yoked control design. To determine whether or not the temporal relationship between the event and the response is relevant to the observed effect, independent groups may be trained with different intervals of delay between response and event. To determine whether the frequency of events is responsible for the observed effects, independent groups may be used with different event frequencies. These may be less efficient than the yoked control design, but the random errors do not produce systematic effects.

Imagine an experiment in which two observers face a screen. On each trial the experimental *S* depresses a key that projects a word on the screen. When he believes he has correctly identified the word, he releases the key which turns off the stimulus word. The matched control *S* sits by his side as a passive observer and receives exactly the same duration and intensity of the stimulus on each trial. The research hypothesis might be that the recognition threshold is decreased if *S* makes a motor response terminating the stimulus to be identified. If this experiment should demonstrate that the experimental *S*s correctly identify a significantly greater proportion of the words than the control *S*s, we would certainly search for an explanation more compelling than the original research hypothesis. Such an explanation would be found in the systematic effect of individual differences between *S*s and the moment-to-moment variability of each *S* in the probability of detection of a stimulus.

If the yoked control design had been used primarily to investigate such unlikely research hypotheses, it would have been severely criticized and abandoned some time ago. The design, however, typically has been used to test research hypotheses which are probably correct. For this reason the yoked control design may not often have led to the false rejection of the null hypothesis. Nonetheless, since the design confounds various sources of random error with treatment effects, the results are necessarily ambiguous.

REFERENCES

Bindra, D., & Anchel, H. Immobility as an avoidance response and its disruption by drugs. *J. exp. Anal. Behav.*, 1963, **6**, 213–218.

Brady, J. V., Porter, R. W., Conrad, D. G., & Mason, J. W. Avoidance behavior and the development of gastroduodenal ulcers. *J. exp. Anal. Behav.*, 1958, **1**, 69–72.

Brimer, C. J., & Kamin, L. J. Fear of the CS in avoidance training and fear from a sense of helplessness. *Canad. J. Psychol.*, 1963, **17**, 188–193.

Carnathan, J., & Church, R. M. The effect of competitive allocation of reinforcements to rats in a straight alley. *J. gen. Psychol.*, 1964, in press.

Church, R. M. Effects of a competitive situation on the speed of response. *J. comp. physiol. Psychol.*, 1961, **54**, 162–166.

Church, R. M. Effect of relative skill on the amount of competitive facilitation. *Psychol. Rep.*, 1962, **11**, 603–614.

Church, R. M. The varied effects of punishment on behavior. *Psychol. Rev.*, 1963, **70**, 369–402.

Church, R. M., & Carnathan, J. Differential reinforcement of short-latency responses in the white rat. *J. comp. physiol. Psychol.*, 1963, **56**, 120–123.

Clayton, Frances L., & Savin, H. B. Strength of a secondary reinforcer following continuous or variable ratio primary reinforcement. *Psychol. Rep.*, 1960, **6**, 99–106.

Crowder, W. F., Gay, B. R., Bright, M. G., & Lee, M. F. Secondary reinforcement or response facilitation? III. Reconditioning. *J. Psychol.*, 1959, **48**, 307–310.

Crowder, W. F., Gay, B. R., Fleming, W. C., & Hurst, R. W. Secondary reinforcement or response facilitation? IV. The retention method. *J. Psychol.*, 1959, **48**, 311–314.

Crowder, W. F., Gill, K., Jr., Hodge, C. C., & Nash, F. A., Jr. Secondary reinforcement or response facilitation? II. Response acquisition. *J. Psychol.*, 1959, **48**, 303–306.

Crowder, W. F., Morris, J. B., & McDaniel, M. H. Secondary reinforcement or response

facilitation? I. Resistance to extinction. *J. Psychol.*, 1959, **48**, 299–302.

Ferster, C. B., & Skinner, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

Fox, R. E., & King, R. A. The effects of reinforcement scheduling on the strength of a secondary reinforcer. *J. comp. physiol. Psychol.*, 1961, **54**, 266–269.

Gibson, Janice. Effects of noxious stimulation whose termination is contingent on a response and of noxious stimulation whose termination is arbitrary. Unpublished master's thesis, Brown University, 1957.

Kimble, G. A., Mann, Lucie I., & Dufort, R. H. Classical and instrumental eyelid conditioning. *J. exp. Psychol.*, 1955, **49**, 407–417.

Kling, J. W., Horowitz, L,. & Delhagen, J. E. Light as a positive reinforcer for rat responding. *Psychol. Rep.*, 1956, **2**, 337–340.

Logan, F. A. A comparison of avoidance and nonavoidance eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 391–393.

Logan, F. A. *Incentive.* New Haven: Yale Univer. Press, 1960.

Moore, J. W., & Gormezano, I. Yoked comparisons of instrumental and classical eyelid conditioning. *J. exp. Psychol.*, 1961, **62**, 552–559.

Mowrer, O. H., & Viek, P. An experimental analogue of fear from a sense of helplessness. *J. abnorm. soc. Psychol.*, 1948, **43**, 193–200.

Sawrey, W. L., Conger, J. J., & Turrell, E. An experimental investigation of the role of psychological factors in the production of gastric ulcers in rats. *J. comp. physiol. Psychol.*, 1956, **49**, 457–461.

Sawrey, W. L., & Weisz, J. D. An experimental method of producing gastric ulcers. *J. comp. physiol. Psychol.*, 1956, **49**, 269–270.

Shearn, D. W. Operant conditioning of heart rate. *Science,* 1962, **137**, 530–531.

Stanley, W. C., & Monkman, J. A. A test for specific and general behavior effects of infantile stimulation with shock in the mouse. *J. abnorm. soc. Psychol.*, 1956, **53**, 19–22.

Wyckoff, L. B., Sidowski, J., & Chambliss, D. J. An experimental study of the relationship between secondary reinforcing and cue effects of a stimulus. *J. comp. physiol. Psychol.*, 1958, **51**, 103–109.

Zimmerman, D. W. Sustained performance in rats based on secondary reinforcement. *J. comp. physiol. Psychol.*, 1959, **52**, 353–358.

(Received September 26, 1963)

Psychological Bulletin
1964, Vol. 62, No. 2, 132–144

RELATIONSHIP BETWEEN EEG AND TEST INTELLIGENCE: A CRITICAL REVIEW [1]

WILLIAM VOGEL AND DONALD M. BROVERMAN
Worcester State Hospital, Massachusetts

Literature on the relationship of EEG to test intelligence scores is reviewed. Evidence for the relationship between these 2 variables seems strongest for samples of children, institutionalized geriatric patients, mental deficients, and brain-injured persons; and weakest for samples of normal adults. EEG indices seem to be more strongly related to MA than to IQ. Criticisms of the previous research are directed at insufficient attention to the measurement of intelligence, failure to control for sex, too restricted placements of leads, too restricted an employment of EEG indices, and the assessment of EEG performance under inappropriate conditions. These methodological and theoretical problems are discussed and some solutions proposed.

In an extensive review of the psychological correlates of EEG, Lindsley (1944) concluded that "it appears doubtful that there is any very high degree of relationship between intelligence as measured by tests and the EEG." Ostow (1950) stated that "there is no relation between any feature of the encephalogram and the intelligence of nondefective persons." Finally, Ellingson (1956) in the last major review of this topic, wrote "the weight of the evidence indicates that the alpha rhythm is unrelated to test intelligence."

Thus, the virtually unanimous opinion of these reviewers is that little convincing evidence exists for relationships between EEG and test intelligence. However, the total number of studies reviewed by Lindsley (1944), Ostow (1950), and Ellingson (1956) were almost equaled in number by the studies which were not reviewed. In addition, since Ellingson's (1956) paper, as many studies have been reported on the relationship between EEG and test intelligence as were published previously. Since a number of the recent studies as well as many of the older, previously unreviewed papers report significant relationships between EEG and test intelligence, it seems appropriate at this time to evaluate anew the relationship between these two variables.

The relationship of EEG to test intelligence involves the question of the nature of EEG phenomena, and the level of human behavior to which they are related. Electroencephalographers are sharply divided on this issue. On one side are those represented by Sisson and Ellingson (1955) who state:

> Since both alpha and beta activity appear to be quite primitive functions of neural tissue, we find it difficult to believe that any of their measures will be found to correlate with any of the dimensions of so complex and phylogenetically recent an entity as the human personality [p. 357].

Presumably, they would hold that this view is also applicable to the attempt to correlate EEG with intelligence. In

[1] This work was supported by the Dementia Praecox Research Project, Worcester State Hospital and Research Grants M-5773 and M-6369, from the National Institute of Mental Health, United States Public Health Service. The paper was prepared while the senior author was a United States Public Health Service Postdoctoral Clinical Fellow at Worcester State Hospital, Massachusetts.

support of this argument it might be noted that alpha, and alpha blocking by light stimuli have been demonstrated in animals as phylogenetically primitive as the insect; alpha frequencies have been observed in the sympathetic nervous system of the cat (Bronk, Ferguson, Margarea, & Solandt, 1936), and Darrow (1950) argues powerfully for subcortical influences upon the human cortical alpha rhythm.

Other researchers, on the other hand, expect EEG indices to be related to higher mental processes. There is no reason to assume that the bioelectrochemical events which produce alpha in the insect, in the sympathetic nervous system of the cat, and on the cortex of man have the same significance for mental functioning. It may be that EEG waves have different neural implications depending upon the kind of nervous cell in which they appear: in other words, brain waves in higher neural centers may reflect activity of a more advanced biological process than similar waves which appear in lower neural centers. In this view, brain waves recorded from higher level tissue may be related to the psychological functions which are vested in that tissue. If this argument is reasonable, and we believe it to be, then it is also reasonable to expect that cortical alpha and beta waves may be related to human intelligence.

Relationships between EEG and Test Intelligence in Mental Defectives

Relationships between EEG indices and intelligence in mental defectives were first reported by Berger (1933) and later by Kreezer (1937, 1938, 1939a, 1939b) and Kreezer and Smith (1936, 1937). These early pioneering efforts were primarily concerned with the relationship between mental age (MA) and various properties of the alpha rhythm, a problem which continues to be a focus of research effort. In one of the more statistically and methodologically sound of these studies Kreezer (1936b) found significant positive relationships between MA and occipital alpha index [2] in Mongolian idiots, and in phenyl pyruvic amentias. A positive relationship between alpha frequency and MA was also found in familial defectives. A later paper (Kreezer & Smith, 1950) virtually replicated the relationship between alpha frequency and MA in familial defectives ($r = .323$, $p < .06$, $df = 34$). In general, these findings are in agreement with the other reports by Berger, Kreezer, and Kreezer and Smith cited above.

Recently a number of European studies have confirmed Berger's and Kreezer's findings of relationships between defective intelligence and alpha frequency. A review by Netchine (1959) referred to studies by Bernhard and Skoglund (1939), who used imbeciles as subjects, and Gunnarson (1945), who studied Mongolian idiots. Both of these reports concluded that mental defectives show a slower alpha frequency than normal subjects of the same age. Novikova (1956) compared normal subjects, from 9 to 16 years old, with

[2] Standard EEG terminology is employed throughout this paper. Thus "delta frequency" refers to any rhythm of 3 cps or slower; "theta frequency," 4–7 cps; "alpha frequency," 8–13 cps, and "beta frequency," 18–30 cps. The term "frequency" is also ordinarily used in the sense of comparing rhythms within a frequency band: for example, a person whose alpha frequency is nearer to 8 than to 13 cps is said to evidence a slow alpha frequency. The term "index" (delta, theta, alpha, or beta index) refers to the percentage of time a particular frequency is present over the entire course of an EEG record. The term "amplitude" refers to the height of the EEG waves of any given frequency.

feebleminded subjects of the same chronological age (CA) who were graded down to the lowest mental levels. She found that slower alpha frequencies are associated with lower MAs at the lower mental levels.

Reports issuing from other laboratories, however, have generally not confirmed Berger's and Kreezer's findings of relationships between alpha index and mental defective intelligence. Rahm and Williams (1938) reported an absence of significant relationships between occipital alpha index and MA in a group of unselected feebleminded adults ranging from 1 to 9.8 years MA. Their results may be attributed to the heterogeneous nature of their sample in regard to kind and degree of mental deficiency, age, sex, etc. Netchine (1959), citing a finding by Corriol and Cain (1949), stated that they did not observe any relationships between MA and alpha index (in a group of feebleminded children) except before the period of puberty.

Finally, Lindsley (1938) reported no significant differences in regard to either occipital alpha index or occipital alpha frequency between mental defectives and normals matched for age, but the number of subjects involved was small (four deficients and four normals).

The above studies focused principally on the relationship between intelligence and EEG indices derived from occipital recordings. In a highly suggestive study, Netchine, Talan, Lairy, and Zazzo (1959) demonstrated that although the correlations between MA and occipital, parietal, rolandic, and frontal alpha frequency were all significant, occipital alpha frequency provided lower correlations with MA than did measures of alpha frequency which were obtained from the parietal, rolandic, or frontal leads. Unfortunately, the conclusiveness of the Netchine et al. study is uncertain, since no information was given in regard to the steps taken to control for chronological age: in the age range sampled in their study (14–28 years), the effect of age upon EEG is commonly conceded to be considerable.

Netchine et al. (1959) and Netchine and Netchine (1962) have attempted to break with the traditional dependence upon frequency bands and to derive novel EEG measures which quantify complex EEG relationships from various regions of the brain. Netchine et al. (1959) reported that one complex EEG pattern which they were able to quantify (based on alpha index and frequency-band scatter) characterized 50% of low-grade feebleminded subjects and 0% of high-grade feebleminded subjects. Netchine and Netchine (1962), employing another EEG index based upon alpha rhythm and an amplitude measure, again successfully differentiated high- and low-grade feebleminded subjects, and also differentiated feebleminded subjects who showed eventual improvement in mental ability from feebleminded subjects who did not.

Two studies have investigated the relationship of defective intelligence to clinically diagnosed EEG abnormalities rather than to indices such as alpha frequency or alpha index. Gibbs, Rich, Fois, and Gibbs (1960) studied 1,118 retarded persons with an IQ of 80 or less, none of whom had epilepsy or cerebral palsy. They reported that the incidence of EEG abnormality was highest in those whose IQ was 60 or less. In a study comparing the EEGs of high- and low-grade mental deficients, Beckett, Bickford, and Keith (1956) found that low-grade deficients evidenced more records with dysrhythmia under rest conditions, and more records with a paroxysmal response to photic stimulation. However, the data presented by

Beckett et al. seem to fall short of statistical significance.

The weight of the evidence, then, supports the proposition that a significant relationship exists between intelligence and various aspects of EEG functioning (especially occipital alpha frequency) among feebleminded persons, particularly at the lower ranges of feeblemindedness. The findings of Netchine et al. (1959) and Netchine and Netchine (1962) suggest that highly significant relationships might well be obtained throughout the range of feeblemindedness by employing electrodes from areas other than the occipital, and by using overall indices which combine measures of frequency, amplitude, and patterning.

Relationships between EEG and Test Intelligence in Children

The studies cited in this section employed children of normal intelligence although Knott, Friedman, and Bardsley (1942); Aird and Cohen (1950); and Netchine and Lairy (1960) also included small samples of mentally deficient children in order to study the full range of intelligence.

Three studies have reported on the relationship between occipital alpha frequency and intelligence in children. There are wide differences between the studies in regard to age, range of intelligence, and presence or absence of neuropsychiatric complaints. These sample differences may account for the lack of agreement in results, illustrated below.

A significant positive relationship between occipital alpha frequency and intelligence was obtained in a group of 8-year-old children whose IQs ranged from 30 to 171, but not in a group of 12-year-olds whose IQs ranged from 56 to 153 (Knott, Friedman, & Bardsley, 1942). Both positive and negative significant relationships between alpha frequency and intelligence have been reported by Netchine and Lairy (1960) in a sample of 209 children who presented neurological complaints, but who were found to be free of brain damage or serious emotional disorders, and who ranged in intelligence from defective through superior. The correlation between occipital alpha frequency and intelligence at age 5–6 years was −.34; at age 7–8 years, .20; at age 9–10, .30; at age 11–12, .38 (all correlations, except that at age 7–8 years, are statistically significant). Thus, the relationship between intelligence and occipital frequency was significantly negative at the younger age levels and then became increasingly positive as age increased. Finally, no statistically significant relationships between intelligence and occipital alpha frequency in children of normal intelligence who ranged from 5 to 11 years of age, were found at any 1-year age level by Henry (1944).

Lindsley (1938) reported no significant relationships between occipital alpha index or pattern of occipital alpha waves with intelligence, in normal children 10–14 years of age.

Two studies reported significantly negative relationships between slow rhythms (delta or theta) and intelligence in children. In a study of the relationship between intelligence and delta index from both central and occipital areas in children of normal intelligence who ranged from 8 to 11 years of age, Henry (1944) found significant inverse correlations between amounts of precentral delta and intelligence at ages 9 and 10, and between amount of occipital delta and intelligence at age 9. The data of Netchine and Lairy (1960) strongly suggested negative relationships between intelligence and the amount of theta present,

at each 1-year age level from 5 through 12 years.

The degree of EEG abnormality in cerebral palsied "children" ranging from 1 to 21 years in age, and in intellectual level from the lowest grades of feeble-mindedness through the superior range, was investigated in a study by Aird and Cohen (1962). They reported no significant relationship between intelligence and EEG abnormality in the overall sample. It is unclear from their report how intelligence was evaluated at the younger age levels.

It is noteworthy that each of the three studies reporting statistically significant relationships between EEG and test intelligence (Henry, 1944; Knott et al., 1942; Netchine & Lairy, 1960) used samples at different age levels, but held age within each sample constant. On the other hand, the two studies that did not report significant relationships employed subject samples that were comparatively heterogeneous in regard to age. Lindsley's (1938) subjects, for instance, ranged from 10 to 14 years, while the subjects in Aird and Cohen's (1950) study ranged in age from 1 to 21 years. This strongly suggests that EEG-test intelligence relationships in children are obscured if CA is not held constant. When CA is held constant, however, IQ becomes perfectly correlated to MA, since $IQ = MA/CA \times 100$. Thus, it would seem that EEG is related to MA rather than to IQ in children. In this case, the results of the studies with children are consistent with those on mental defectives; that is, relationships between EEG and MA were found in both sets of studies.

Theoretically, it seems plausible that MA, more than IQ, should relate to EEG functioning, since MA reflects the absolute level of mental ability of an individual, whereas IQ reflects the level of mental ability of an individual relative to his chronological age. In this sense, MA is a better reflection of the absolute level of mental activity (and presumably brain functioning) than is IQ.

Relationships between EEG and Test Intelligence in Normal Adults

Six studies have reported on the relationships between alpha frequency and intelligence in adults. Of these, three found no significant relationships (Biesheuvel & Pitt, 1955; Gastaut, 1960; Shagass, 1946); two studies (Mundy-Castle, 1958; Mundy-Castle & Nelson, 1960) reported positive, significant relationships; and one study reported a significant negative relationship (Sugarman, 1961). Except for the Mundy-Castle and Nelson (1960) study which employed adult subjects of whom half had less than normal intelligence, these studies employed adult subjects of at least average mental ability.

This vast disparity in the results of studies of the relationship between alpha frequency and adult intelligence poses obvious problems. Mundy-Castle (1958) has provided a possible solution in his discussion contrasting findings of significant positive relationships between alpha frequency and adult intelligence to the lack of relationship found between these variables by Shagass (1946) and Biesheuvel and Pitt (1955). He noted that he alone used the Wechsler-Bellevue test to measure intelligence. The correlations between the Wechsler-Bellevue and other measures of normal adult intelligence (Wechsler, 1944) indicate considerable variance unique to each type of intelligence test. Therefore, alpha frequency may be significantly correlated with the Wechsler, and the Wechsler significantly correlated with the other intelligence tests, without alpha frequency being correlated with tests of intelligence other than the Wechsler. In sum, Mundy-Castle's (1958) argument is that alpha

frequency in normal adults is differentially correlated with particular aspects of intellectual functioning, which, in turn, are differentially measured by various intelligence tests. Of the three studies reported since Mundy-Castle's (1958) argument was presented (Gastaut, 1960; Mundy-Castle & Nelson, 1960; Sugarman, 1961) only Mundy-Castle and Nelson employed the Wechsler, and once more a significant positive relationship between alpha frequency and total intelligence test score was obtained. Mundy-Castle (1958) believes that such relationships may be attributable to nonintellective (or tempermental) electrophysiological components of intelligence.

In contrast with the wide range of results obtained with alpha frequency, no studies reported significant relationships between test intelligence and alpha index (Gastaut, 1960; Mundy-Castle, 1958; Sugarman, 1961) or between test intelligence and alpha amplitude (Gastaut, 1960; Sugarman, 1961).

Relationships between EEG and Test Intelligence in Older Persons

In a comprehensive review of the literature on EEG in old age, Obrist and Busse (1962) noted relationships between EEG slowing and intellectual deterioration. However, the majority of the studies reviewed by Obrist and Busse employed psychiatric judgments or rating scales to categorize intellectual functioning. Only those studies which employed standard psychometric tests will be discussed here.

Studies which employed samples consisting in large part or entirely of hospitalized geriatric men and women indicate that subjects with normal EEGs evidence significantly higher intelligence test scores than subjects of the same CA with abnormal slow EEGs (Barnes, Busse, & Friedman, 1956; Silverman, Busse, Barnes, Frost, & Thaler, 1953). On the other hand, a study of nonhospitalized geriatric men and women (Busse, Barnes, Friedman, & Kelty, 1956) found no IQ differences between groups of subjects of the same mean age with normal versus abnormal (focal, slow, or mixed) EEGs. Similarly, Thaler (1956) found that the age-corrected Wechsler-Bellevue Form I (WBI) IQs of nonhospitalized, aged men and women with normal and focal EEGs were not significantly different from similar subjects with diffuse slow and mixed EEGs. However, although the four EEG groups in Thaler's study did not differ in age, the total weighted scores, uncorrected for age, of the normal and focal EEG groups were significantly higher than those of the slow and mixed EEG groups.

The hypothesis that different EEG-intelligence test relationships are obtained in geriatric samples depending upon whether hospitalized or nonhospitalized subjects are employed was tested by Obrist, Busse, Eisdorfer, and Kleemeier (1962). In two institutionalized samples, but not in a noninstitutionalized sample, subjects with diffuse slow activity evidenced significantly lower Full Scale, Verbal, and Performance raw scores on either the WBI or the Wechsler Adult Intelligence Scale (WAIS) than did subjects with normal EEGs. However, since there was significant variation in the ages of several of the institutionalized subgroups, Obrist et al. concluded that their positive results might be attributable to age, independent of EEG type.

The relationship between alpha frequency and test scores was also investigated by these authors. One institutionalized sample evidenced significant, positive correlations between alpha frequency and Verbal, Performance, and Full Scale row scores; and the other institutionalized group evidenced positive correlations between alpha fre-

quency and Performance raw scores. No significant relationships were found in the noninstitutionalized group. When a correction was introduced for age and arteriosclerosis, it was found that the positive relationship between alpha frequency and IQ existed only in a hospitalized arteriosclerotic sample. Consequently Obrist et al. concluded that the relationships between EEG indices and intelligence in their geriatric samples were attributable to variations in health status of the subjects. In a review article Obrist and Busse (1962) cite further evidence for such a view; however, Thaler (1956) found relationships between EEG and intelligence in a normal group of geriatric subjects.

Despite an extensive literature on the EEGs of geriatric subjects no one has yet reported an investigation relating EEG to intelligence in a geriatric sample with adequate controls for age, health, and estimated pregeriatric intellectual level. Such a study is badly needed, in part to control for individual differences in aging. In addition, although each of the above studies included both male and female subjects, none of the studies investigated the effect of sex on the correlation between EEG and intelligence.

The same disparity between mental ability and IQ previously observed in the studies of children is also present in the geriatric studies. When intelligence test scores uncorrected for age are used as intellectual indices, absolute mental ability is being assessed; however, when test scores are age corrected, IQ is being assessed. The two phenomena are not comparable. As with children and feebleminded persons, mental ability seems to be more closely related to the EEGs of geriatric persons than is IQ.

The studies of geriatric patients have frequently confounded mental ability and CA, as have some studies of children. A negative correlation between CA and mental ability usually exists in geriatric persons. Hence, when a geriatric sample includes subjects whose ages span a number of years, CA and mental ability are confounded. To control for this effect, geriatric studies might compare samples of homogeneous age (persons 60–61 years of age, 64–65 years, etc.) rather than just compare samples of similar mean age. Relationships between test intelligence and EEG would then be solely a function of mental ability.

Relationships between EEG and Test Intelligence in Persons with Organic Brain Disease

The relationships between EEG disturbance and intellectual disturbance in persons with organic brain disease has been investigated by Greenblatt, Levine, and Atwell (1945); Hoedmaker and Murray (1952); and Silverman and Harris (1954). Each study reports significant positive relationships. However, these studies will not be evaluated here since in each case the index of intellectual disturbance was based on a complex judgment of disturbance of cognitive functioning, rather than on just standardized intelligence test performances.

Two studies have been reported in which EEG disturbance in patients with organic brain damage has been related to performance on the Ravens Progressive Matrices. A Ravens mean IQ of 100 in epileptics without diffuse dysrhythmia, as opposed to a mean IQ of 89 in subjects with diffuse dysrhythmia, has been reported by Vislie and Hendriksen (1958). Hill (1952) reports that Rey, using the Progressive Matrices, the Mill Hill Vocabulary, and the Wechsler Bellevue, found that brain-injured persons with paroxysmal EEGs had lower Verbal than Performance scores, unlike brain-injured patients with nonparoxysmal EEGs.

Strong evidence for the existence of

EEG correlates of specific mental abilities is provided in a study by Klove (1959), who found that brain-damaged patients with EEG disturbance over the right hemisphere had significantly higher WBI Verbal than Performance scores. Brain-damaged patients with EEG disturbance over the left hemisphere had significantly higher Performance than Verbal scores. In addition, although there were no differences in the Full Scale raw scores, the left hemisphere disturbance group had significantly higher Performance raw scores than the right hemisphere disturbance group, but significantly lower Verbal raw scores. The two groups had been roughly equated for age, education, and sexual composition.

Kooi and Hovey (1957) report a study in which epileptics were simultaneously administered intelligence tests and EEGs; disturbances in higher integrative mental processes were significantly associated with concomitant paroxysmal activity. Earlier (1955) they reported associations between paroxysmal EEGs and particular intelligence test abnormalities.

Studies of Frequency Pattern Variability

Studies of the relationship between intelligence and "frequency pattern variability" (i.e., the range of frequencies evidenced in the EEG record of an individual in a given time sample) have produced markedly conflicting results. Walter (1953) suggested that intelligent persons would show greater variability than unintelligent persons, but produced little evidence to substantiate his hypothesis. Ellingson, Wilcott, Sineps, and Dudek (1957) studied the EEG records of 27 male and female adults and found no correlation between frequency-pattern variability and IQ. However, Netchine, Talan, Lairy, and Zazzo (1959) found a high negative correlation (−.76) between IQ and frequency-pattern variability in a group of 30 adolescents and adults (aged 14–28 years) who were below normal intelligence. Given the ages of their subjects, however, it seems likely that Netchine et al. may have capitalized on EEG variability that might be as justifiably attributed to CA as to intelligence.

Discussion

Two major conclusions may be drawn from this review of EEG-test intelligence relationships. First, the bulk of the studies with feebleminded subjects, children, institutionalized geriatric subjects, and brain-injured adults have reported significant EEG-test intelligence relationships. However, with the exception of Mundy-Castle (1958) and Mundy-Castle and Nelson (1960), the investigators who have studied normal adults have not found significant relationships between test intelligence and EEG tracings. It appears, then, that relationships between EEG and test intelligence are most evident in subjects who have either relatively undeveloped intellectual function (i.e., children; feeble-minded subjects) or deteriorated intellectual function (brain-damaged and institutionalized geriatric subjects).

Second, it seems likely that EEG tracings in children, feebleminded, and geriatric subjects are related to absolute mental ability, or MA, rather than IQ.

That so many significant findings have been reported is, in a sense, surprising, since a number of methodological problems exist which, potentially, could obscure existing intelligence-EEG relationships. Each of these methodological issues will be discussed below.

The Measurement of Intelligence. The reviewed studies indicate that relatively more attention has been accorded to problems in the measurement of EEG than to issues involved in the measurement of intelligence. While it is com-

monly assumed that mental abilities reach their peak and then plateau in early adulthood, evidence exists that increments in absolute levels of mental ability continue for a considerable period beyond early adulthood. Bayley (1955), for example, has argued that if appropriate measuring instruments are used, the continuous development of absolute levels of mental ability can be demonstrated throughout the first 50 years of life. Christian and Paterson (1936) have shown that particular verbal skills continue to increase up to age 60 or even 70. To the extent that this view of intellectual development is correct it becomes increasingly important to obtain subjects of the same age whenever intelligence is studied in adult and elderly subjects, as well as in children. It is worth noting that in the studies in which the relationship between EEG and test intelligence is most ambiguous, that is, those studies employing normal adults, no attempt was made to experimentally vary age, and in only two of the six studies was an attempt made to control for possible effects of age (Gastaut, 1960; Mundy-Castle & Nelson, 1960). On the other hand, where age has generally been controlled or experimentally varied (i.e., in studies of children, mental deficients, and to a lesser extent impaired geriatric subjects), the nature of the relationship between EEG and test intelligence is relatively clearer.

The work of Garrett (1946), Guilford (1959), and Werner (1948) on the development of intelligence may additionally clarify the results obtained in investigations of the relationship between EEG and test intelligence. Garrett (1946) has demonstrated that whereas intelligence is a relatively unitary phenomenon in young children, it becomes increasingly differentiated during adolescence into a number of relatively independent factors. Guilford's work suggests that the differentiation of intelligence continues into adulthood. He reports that 50 intellectual factors are currently known in adult intelligence, and argues that as many as 70 more may be isolated in the future. Garrett's and Guilford's results are compatible with Werner's (1948) who posits that a salient aspect of development is increasing differentiation. Werner, in addition, regards feebleminded adults, impaired aged persons, brain-damaged persons, and children as evidencing a structurally similar lack of differentiation of cognitive processes.

If Garrett, Guilford, and Werner are correct, then any given test should be a fairly accurate index of intelligence for persons who evidence a lack of differentiation of cognitive faculties (children; feebleminded, brain-damaged, or impaired geriatric subjects). These subjects should perform similarly across a wide range of tests, thus increasing the likelihood of a significant correlation between any given measure of intelligence, and any other variable which may be related to intelligence (such as EEG). On the other hand, a given test would reflect only a very narrow band of the total intellectual spectrum in a cognitively differentiated adult. A normal adult would score very differently on particular tests depending upon which aspects of intelligence were measured by the tests, and depending upon his special mental skills. Consequently, the correlations of any particular measure of adult intelligence with any other variable (such as EEG) would show wide variation across individual adults and individual tests. In short, this argument supports Mundy-Castle's explanation of the wide disparity of results obtained in studies of EEG and test intelligence relationships in adults, when different intelligence tests are used.

Placement of Leads. Most of the studies reviewed here have used only occipital leads. However, while numerous theorists have thought that higher

intellectual functions in humans are localized in the frontal, temporal, or parietal lobes, few have argued that the occipital lobes play a dominant role in intellectual functioning. It is true that Lashley (1929) demonstrated, at least in lower animals, that all parts of the brain are important in intellectual functioning ("mass action"); however, Werner (1948) argues that mass action is an attribute of low phylogenetic and ontogenetic development. Weinstein and Teuber's work (1957) suggests that the occipital lobes in adults are relatively uninvolved in intellectual functioning, at least as measured by one standard intelligence test. Testing healthy, normal soldiers prior to subsequent brain injury, and again after recovery from injury, and comparing these "before and after brain injury" measurements with test measurements on a normal control group, Weinstein and Teuber concluded that left parietal or temporal injury leads to the greatest intellectual loss, and frontal and occipital injury the least intellectual loss. Weinstein and Teuber's conclusion is supported in part by the work of Netchine et al. (1959) who, as we have seen, found that the highest correlations appeared between nonoccipital EEG measurements and intelligence, and that the correlations between occipital EEG measurements and IQ were the lowest of all. In this regard it is noteworthy that occipital EEG has been most successfully related to intelligence in those classes of subjects whose cognitive functions (and presumably whose brain function) were relatively undifferentiated, at least from the theoretical point of view of Werner (children; feebleminded, brain-damaged, and impaired geriatric subjects).

Conditions of EEG Recording. Each of the studies reviewed investigated relationships between test intelligence and an EEG taken when the subject was physically at rest and mentally inactive, and perhaps also during some other standard EEG conditions. With the exception of the study of Kooi and Hovey (1957), none of the studies employed conditions involving intelligent behavior; that is, none required subjects of varying degrees of intellectual ability to solve intellectual problems while the EEG was being taken. The question may be asked, why should the EEGs of persons differing in intelligence vary under conditions in which nothing is done to stimulate or activate those differences in intelligence? The principle that behavior in any instance is a function of personality traits interacting with stimulus conditions is no longer new in psychology. The same principle is, in fact, well accepted in electroencephalography. It is accepted, for example, that epilepsy may not be revealed in an EEG unless the specific conditions to which the pathological brain tissue is especially sensitive are present; it is for this reason that a myriad of activating techniques have come into use (Schwab, 1951). Unfortunately, this same degree of sophistication has not been applied to investigations of the relationships of EEG tracings to test intelligence.

This is not to argue that the EEGs of persons who differ in intelligence should not differ under rest; there is, as shown above, evidence that they do differ (although the strength of the evidence varies in differently defined subject samples). It is suggested, however, that such EEG differences are more likely to become accentuated and apparent under conditions which require intellectual activity than under conditions which do not encourage active thought. This argument is not entirely new; similar suggestions have been put forth by Knott (1940) and by Lairy (1960).

Sex. Kreezer (1939a) demonstrated significant differences between male and female mental deficients in alpha frequency and alpha amplitude; Obrist

and Busse (1962) and Mundy-Castle (1962) noted that mean alpha frequencies were reliably faster among aged females than among aged males; Obrist and Busse (1962) noted that aged females render more EEG fast activity than aged males; Mundy-Castle (1951) reported that theta and beta were present to a significantly greater extent in young females than in young males; Golla, Hutton, and Walter (1943) reported differences in alpha activity between males and females although they did not offer statistical proof; and Smith (1954) reported significant differences between the EEG records of males and females. The great majority of studies which have investigated EEG-intelligence test relationships employed both males and females in the same subject sample. However, little has been done, excepting Kreezer's (1939a) early investigation, to determine how sex differences affect the correlation between EEG and test intelligence: Kreezer's work showed marked differences in the relationship between EEG and test intelligence in female compared to male mental deficients.

EEG Indices. The American studies reviewed here have generally limited their investigations of test intelligence-EEG relationships to some aspect of the (occipital) alpha rhythm, although slow waves (delta, theta) have also been studied, especially by those who have worked with geriatric samples. That this concern with the traditional frequency bands might not be the only appropriate approach to this problem was first recognized by Knott (1940), and later, Brazier (1952, 1953). Subsequently, some European investigators have utilized complex EEG indices which are composites of different aspects of the EEG (Netchine & Lairy, 1960; Netchine & Netchine, 1962; Netchine et al., 1959; Walter, 1953). The results of these studies indicate that the use of such indices is a most promising methodology in investigations of the relationship of EEG and test intelligence. In addition it would seem worthwhile to relate intelligence to aspects of kappa waves, which appear to be temporal EEG waves which are prominent during intellectual activity. The literature on kappa through 1955 has been reviewed by Ellingson (1956).

With the exception of the Sugarman (1961) and Netchine and Lairy (1960) studies, in every case in which test intelligence has been found to be related to EEG frequencies, low intelligence was associated with slow alpha frequencies and the presence of the slower EEG rhythms (delta and theta). Conversely, higher levels of intelligence were found associated with the faster alpha frequencies and an absence of the slow delta and theta rhythms. This suggests that attention might profitably be directed at the still faster beta waves, which typically occur while the organism is mentally or physically active. In any case, there is considerable evidence that electrocortical activity and mental activity are positively correlated. The hypothesis deserves further scrutiny that the EEG may be related to a physiological process underlying human intelligence.

REFERENCES

AIRD, R. B., & COHEN, R. EEG in cerebral palsy. *J. Pediat.,* 1950, **37**, 448–454.

BARNES, R. H., BUSSE, E. W., & FRIEDMAN, E. L. The psychological functioning of aged individuals with normal and abnormal electroencephalograms: II. A study of hospitalized individuals. *J. nerv. ment. Dis.,* 1956, **124**, 585–593.

BAYLEY, NANCY. On the growth of intelligence. *Amer. Psychologist,* 1955, **10**, 805–818.

BECKETT, P. G. S., BICKFORD, R. G., & KEITH, H. M. The EEG and various aspects of mental deficiency. *AMA J. Dis. Children,* 1956, **92**, 374–381.

BERGER, H. Uber das Electroenkephalogram des Menschen. *Arch. Psychiat. Nervenkr.,* 1933, **98**, 231–254.

BERNHARD, C. G., & SKOGLUND, C. R. Recherches sur la fréquence alpha de l'electro-

encepholagramme chez l'enfant. *Acta psychiat., Copenhagen,* 1939, **14**, 223–231.

BIESHEUVEL, S., & PITT, D. R. Some tests of speed and tempo of behavior as predictors of the primary-secondary function temperament variable. *J. Nat. Inst. Personnel Res.,* 1955, **6**, 95–102.

BRAZIER, MARY A. Electroencephalography. In E. A. Spiegel (Ed.), *Progress in neurology and psychiatry.* Vol. 7. New York: Grune & Stratton, 1952. P. 262.

BRAZIER, MARY A. Electroencephalography. In E. A. Spiegel (Ed.), *Progress in neurology and psychiatry.* Vol. 8. New York: Grune & Stratton, 1953. P. 257.

BRONK, D. W., FERGUSON, L. K., MARGARIA, R., & SOLANDT, D. Y. Activity of cardiac sympathetic centers. *Amer. J. Physiol.,* 1936, **117**, 237–249.

BUSSE, E. W., BARNES, R. H., FRIEDMAN, E. L., & KELTY, E. J. Psychological functioning of aged individuals with normal and abnormal EEGs: I. A study of nonhospitalized community volunteers. *J. nerv. ment. Dis.,* 1956, **124**, 135–141.

CHRISTIAN, ALICE M., & PATERSON, D. G. Growth of vocabulary in maturity. *J. Psychol.,* 1936, **1**, 167–169.

CORRIOL, J. O., & CAIN, J. L'EEG des enfants arriérés et caracteriels. *Sem. Hôp. Paris,* 1949, No. 2735.

DARROW, C. W. A new frontier: Neurophysiological effects of emotion on the brain. In M. L. Reymert (Ed.), *Feelings and emotions.* New York: McGraw-Hill, 1950. Pp. 247–260.

ELLINGSON, R. J. Brain waves and problems of psychology. *Psychol. Bull.,* 1956, **53**, 1–34.

ELLINGSON, R. J., WILCOTT, R. C., SINEPS, J. G., & DUDEK, F. J. EEG frequency-pattern variation and intelligence: A correlational study. *EEG clin. Neurophysiol.,* 1957, **9**, 657–660.

GARRETT, H. E. A developmental theory of intelligence. *Amer. Psychologist,* 1946, **1**, 372–378.

GASTAUT, H. Correlations between the electroencephalographic and the psychometric variables (MMPI, Rosenzweig, Intelligence Tests). *EEG clin. Neurophysiol.,* 1960, **12**, 226–227.

GIBBS, ERNA L., RICH, CATHERINE L., FOIS, A., & GIBBS, F. A. Electroencephalographic study of mentally retarded persons. *Amer. J. ment. Defic.,* 1960, **65**, 236–247.

GOLLA, F. L., HUTTON, E. L., & WALTER, W. G. The objective study of mental imagery. *J. ment. Sci.,* 1943, **89**, 216–223.

GREENBLATT, M., LEVINE, S., & ATWELL, C. Comparative value of EEG and abstraction tests in diagnosis of brain damage. *J. nerv. ment. Dis.,* 1945, **102**, 385.

GUILFORD, J. P. Three faces of intellect. *Amer. Psychologist,* 1959, **14**, 469–479.

GUNNARSON, S. Electroencephalographic examination of imbeciles. *Acta paediat.,* 1945, **32**, 240–248.

HENRY, C. E. Electroencephalograms of normal children. *Monogr. Soc. Res. Child Develpm.,* 1944, 49(Whole No. 39).

HILL, D. EEG in episodic psychotic and psychopathic behavior: A classification of data. *EEG clin. Neurophysiol.,* 1952, **4**, 419–442.

HOEDMAKER, E. D., & MURRAY, M. Psychologic tests in diagnosis of organic brain disease. *Neurology,* 1952, **2**, 144.

HOVEY, W. B., & KOOI, K. A. Transient disturbances of thought processes and epilepsy. *AMA Arch. Neurol. Psychiat.,* 1955, **74**, 287–291.

KLOVE, H. Relationship of differential electroencephalographic patterns to distributions of Wechsler-Bellevue scores. *Neurology,* 1959, **9**, 871–876.

KNOTT, J. R. Comments upon some electroencephalographic correlates of intelligence. *Yearbk. Nat. Soc. Stud. Educ.,* 1940, **39**(1), 133–135.

KNOTT, J. R., FRIEDMAN, H., & BARDSLEY, R. Some electroencephalographic correlates of intelligence in eight-year-old and twelve-year-old children. *J. exp. Psychol.,* 1942, **30**, 380–391.

KOOI, K. A., & HOVEY, W. B. Alterations in mental function and paroxysmal cerebral activity. *AMA Arch. Neurol. Psychiat.,* 1957, **78**, 264–271.

KREEZER, G. The dependence of the electroencephalogram upon intelligence level. *Psychol. Bull.,* 1937, **34**, 769–770.

KREEZER, G. The electroencephalogram and its use in psychology. *Amer. J. Psychol.,* 1938, **51**, 737–759.

KREEZER, G. Intelligence level and occipital alpha rhythm in the Mongolian type of mental deficiency. *Amer. J. Psychol.,* 1939, **52**, 503–532. (a)

KREEZER, G. The relation of intelligence level and the electroencephalogram. *Yearbk. Nat. Soc. Stud. Educ.,* 1939, **39**, 130–133. (b)

KREEZER, G., & SMITH, F. W. Electrical potentials of the brain in certain types of mental deficiency. *Arch. Neurol. Psychiat.,* 1936, **36**, 1206–1213.

KREEZER, G., & SMITH, F. W. Brain potentials in the hereditary types of mental deficiency. *Psychol. Bull.,* 1937, **34**, 535–536.

KREEZER, G., & SMITH, F. W. The relation of the alpha rhythm of the EEG and intelligence level in the nondifferentiated familial

type of mental deficiency. *J. Psychol.*, 1950, **29**, 47–51.

LAIRY, G. C. EEG et neuropsychiatrie infantile. *Psychiat. Enfant*, 1960, **3**(2), 528–608.

LASHLEY, K. S. *Brain mechanisms and intelligence.* Chicago: Univer. Chicago Press, 1929.

LINDSLEY, D. B. Electrical potentials of the brain in children and adults. *J. gen. Psychol.*, 1938, **19**, 285–306.

LINDSLEY, D. B. Electroencephalography. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 2. New York: Ronald Press, 1944. Pp. 1033–1091.

MUNDY-CASTLE, A. C. Theta and beta rhythms in the electroencephalograms of normal adults. *EEG clin. Neurophysiol.*, 1951, **3**, 477–486.

MUNDY-CASTLE, A. C. Electrophysiological correlates of intelligence. *J. Pers.*, 1958, **26**, 184–199.

MUNDY-CASTLE, A. C. Central excitability in the aged. In H. T. Blumenthal (Ed.), *Medical and clinical aspects of aging.* New York: Columbia Univer. Press, 1962. Pp. 575–595.

MUNDY-CASTLE, A. C., & NELSON, G. K. Intelligence, personality and brain rhythms in a socially isolated community. *Nature, London,* 1960, **185**, 484–485.

NETCHINE, G., & NETCHINE, S. Organisation psychologique et organisation bioélectrique cérébrale dans une population d'arriérés mentaux. *Rev. Psychol. Franç.*, 1962, **7**, 241–258.

NETCHINE, S. EEG et intelligence: Revue générale du probleme. *Année Psychol.*, 1959, **59**, 427–438.

NETCHINE, S., & LAIRY, G. C. Ondes cérébrales et niveau mental: Quelques aspects de l'evolution génétique du tracé EEG suivant le niveau. *Enfance,* 1960, No. 4–5, 427–439.

NETCHINE, S., TALAN, I., LAIRY, G. C., & ZAZZO, R. EEG et niveau mental: Étude d'une population à niveau mental faible. *Année Psychol.*, 1959, **59**, 355–373.

NOVIKOVA, L. A. Étude de l'activité électrique du cortex cérébral des enfants oligophrènes. In, *Problemes de l'activité nerveuse superieure des enfants normaux et anormaux.* Moscow: Luria, 1956.

OBRIST, W. D., & BUSSE, E. W. The electroencephalogram in old age. In W. P. Wilson (Ed.), *Applications of electroencephalography in psychiatry.* Durham: Duke Univer. Press, 1964, in press.

OBRIST, W. D., BUSSE, E. W., EISDORFER, C., & KLEEMIER, R. M. Relation of the electroencephalogram to intellectual function in senescence. *J. Gerontol.*, 1962, **17**(2), 197–206.

OSTOW, M. Psychic function and the electroencephalogram. *Arch. Neurol. Psychiat.*, 1950, **64**, 385–400.

RAHM, W. E., & WILLIAMS, A. C. Aspects of the electroencephalogram in epilepsy and feeblemindedness. *Psychiat. Quart.*, 1938, **12**, 230–235.

SCHWAB, R. S. *Electroencephalography in clinical practice.* Philadelphia, Pa.: Saunders, 1951.

SHAGASS, C. An attempt to correlate the occipital alpha frequency of the electroencephalogram with performance on a mental ability test. *J. exp. Psychol.*, 1946, **36**, 88–92.

SILVERMAN, A. J., BUSSE, E. W., BARNES, R. H., FROST, L. L., & THALER, M. B. Studies on the processes of aging: Part 4. Physiologic influences in psychic functioning in elderly people. *Geriatrics,* 1953, **8**, 370–376.

SILVERMAN, A. J., & HARRIS, V. F. Electroencephalography and psychometric testing in brain damaged patients. *J. nerv. ment. Dis.*, 1954, **120**, 31.

SISSON, B. D., & ELLINGSON, R. J. On the relationship between "normal" EEG patterns and personality variables. *J. nerv. ment. Dis.*, 1955, **121**, 353–357.

SMITH, SHEILA M. Discrimination between electroencephalograph recordings of normal females and normal males. *Ann. Eugen., London,* 1954, **18**, 344–350.

SUGARMAN, L. Alpha rhythm, perception and intelligence. *J. Nat. Inst. Personnel Res.*, 1961, **8**, 170–179.

THALER, M. Relationships among Wechsler, Weigl, Rorschach, EEG findings and abstract-concrete behavior in a group of normal aged subjects. *J. Gerontol.*, 1956, **11**, 404–409.

VISLIE, H., & HENDRIKSEN, G. F. Psychic disturbances in epilepsy. In A. M. Lorentz de Haas (Ed.), *Lectures on epilepsy.* Amsterdam, The Netherlands: Elsevier, 1958. Pp. 29–98.

WALTER, W. G. *The living brain.* New York: Norton, 1953.

WECHSLER, D. *The measurement of adult intelligence.* (3rd ed.) Baltimore, Md.: Williams & Wilkins, 1944.

WEINSTEIN, S., & TEUBER, HANS-LUKAS. Effects of penetrating brain injury on intelligence test scores. *Science,* 1957, **125**, 1036–1037.

WERNER, H. *Comparative psychology of mental development.* New York: Follett, 1948.

(Received October 4, 1963)

VOL. 62, NO. 3 SEPTEMBER 1964

Psychological Bulletin

RESEARCH WITH THE DEAF:

IMPLICATIONS FOR LANGUAGE AND COGNITION [1]

HANS G. FURTH

Catholic University of America
Center for Research in Thinking and Language

Deaf people who are deprived of linguistic experience during the formative years seem to provide a unique opportunity to psychologists concerned with the language-cognition relationship. Empirical studies of deaf people's performance on nonverbal cognitive tasks were reviewed. Deaf were found to perform similarly to hearing persons on tasks where verbal knowledge could have been assumed a priori to benefit the hearing. Such evidence appears to weaken a theoretical position which attributes to language a direct, general, or decisive influence on intellective development. The poorer performance of the deaf on some tasks is parsimoniously attributed to either lack of general experience which is no longer manifest by adulthood or to specific task conditions which favor linguistic habits.

One of the most intriguing problems which falls within the domain of the psychological investigator is the relationship of language to cognitive or intellective development. In general, psychologists have not critically examined this relationship and have been content with speculative or anecdotal evidence. It was James (1890) who remarked that it mattered little in which medium thinking was going on and he quoted biographical recollections as evidence that thought processes were developed in a deaf person before English had been learned. Somewhat later Binet and Simon (1905) made similar inferences from Helen Keller's and Laura Bridgman's outstanding achievements in spite of their being both blind and deaf from early childhood. Reliable empirical studies, however, on the proposition that development of language and of thinking are relatively independent processes were lacking. Under the impact of behaviorism and the modern approaches to verbal learning and verbal mediation on the one hand and the stress on oral language learning of deaf children on the other hand, the opposite proposition has been implied, if not explicitly stated. It is easy to find quotations similar to the following in almost any modern psychological account of cognitive development:

> Skill in concept formation is closely linked to the acquisition of language, particularly to labelling. After he has learned the names or labels applied to objects or events, a child is likely to react in the same way to all stimuli having the same labels. This is known as verbal mediation or mediated generalization.

[1] Writing of this paper was aided in part by Grants 198-T and RD 704 from the United States Vocational Rehabilitation Administration. Thanks are expressed to N. A. Milgram for helpful collaboration in the preparation of this paper and to J. Youniss and R. Dowling for critical reading of the manuscript.

Numerous experiments demonstrate that such mediation is of paramount importance in concept formation, problem solving, thinking and learning [Mussen, 1963, p. 37].

At another point, Mussen, citing as partial evidence research of Soviet psychologists, suggests that language gradually "becomes the most important mediator and regulator of behavior [p. 45]."

With regard to the deaf, the above mentioned observations of James and Binet were actually exceptions to the majority of past and present opinions of lay and professional people in a twofold sense. First, the deaf persons mentioned were certainly not average and were cited more to show what is possible than what commonly happens. Secondly, the authors who quoted them sought empirical illustration for their viewpoint on the relative independence of language and thinking. In contrast to this viewpoint writers on deaf people generally tended to consider these two phenomena quite closely related. The view that the deaf child or adult is cognitively different from and inferior to the hearing person even in nonverbal complex learning tasks because he lacks the free use of language is not only congruent with the common notion of a deaf and dumb person, but seems also in accord with more sophisticated ideas about the deaf. Moreover, it is still common to cite deficient verbal performance of the deaf as evidence for their conceptual deficiency, although it should be rather obvious that the use of verbal tasks with deaf persons is less a test of their intelligence than of their verbal skill.

It seems therefore appropriate to review experimental studies on the nonverbal cognitive functioning of deaf people, both for the general reasons by which psychologists study any special group within human society and for special theoretical implications about language and thinking. As an introduction there follows first some preliminary remarks about the life history of deaf people and classifications of terms relative to language and thinking. The main body of this article reviews studies involving complex learning tasks which can be considered closely related to conceptual ability and intelligence. Afterwards two sections on memory and visual perception complete the survey of nonverbal cognitive studies. Finally, the author proposes a tentative theoretical position about the influence of language on intelligence and considers further implications from the reviewed studies of the deaf as well as additional suggestive evidence from language learning of hearing children.

Some Observations about Deaf People

As the deaf person is perhaps the least known and understood of persons with a physical handicap, it is hardly surprising that little attention has been paid to him by scientists who were expressly concerned with language and cognition. It is, however, the considered judgment of this author that no investigator in this area of interest can afford to neglect the presence of persons in our society who have been minimally exposed to the linguistic environment during their developmental period and who appear to go about the business of living—apart from relatively minor social idiosyncracies—in much the same way as other persons.

Most deaf people have been born to hearing parents. Young deaf children not only do not know the spoken language of their society, but in general, the level of verbal or gestural communication in which they are able to engage with their parents, teachers, and peers is of the most primitive and limited kind. Only after they enter

special schools for the deaf do they receive consistent instruction in language, by reading and writing, and by lip reading and speaking. If one observes these children in the primary grades, one concludes that their ordinary life and play are not strikingly different from that of other children their age. They are apparently normal children growing up in a society and culture which is intimately bound up with language despite the fact that they themselves have had minimal direct exposure to the all-pervading linguistic environment. A number of questions, it would seem, are raised as the behavioral scientist observes these children. How has language deficiency affected the cognitive development of these children? Have they organized their experiences on meaningful perceptual and conceptual bases? Do they manifest the beginnings of logical reasoning? Can they recall things, observe relevant aspects of a whole, and can they form theoretical judgments?

As for the deaf adult, we may notice that they use formal signs, gestures, and manual spelling to converse with each other. A typical male adult may be employed as a semiskilled worker on an assembly line, be married to a deaf wife, and have hearing children. His mastery of English is quite poor, with reading level not above Grade 3; and his speech and lipreading are so limited as to be of little functional value on their own. Note that in contrast to hearing people with a similar reading level, Grade 3 written English constitutes the ceiling of this deaf person's comprehension of language. The conventional sign language, which is customarily not taught in schools for the deaf, is usually acquired during later childhood informally from other deaf people and it is used rather effectively for various social purposes. Deaf adults, who are as thoroughly "at home" in English as any hearing person, either lost their hearing after the establishment of language or do not have so serious a hearing loss as to be justifiably classified among the deaf, or finally, they are rare exceptions.

For all practical purposes, the typical deaf person, whether child or adult, is a language-deficient person both in his present functioning and in his past experience. Even though we may accept the conventional sign language of the deaf as a "true" language and the adult deaf may employ it for everyday communication, it still remains true that only some deaf children of deaf parents—a small minority of all deaf people—learned these signs before age 6. This peculiar status with respect to language makes the deaf a unique experimental subject for clarifying the influence of language on cognition.

Meaning of Terms

Definition of some terms used thus far is already overdue. *Language* refers to the living language as heard and spoken in our society. Knowing a language means mastery of a particular language so that its structure and ordinary vocabulary are implicitly understood and employed by a person. In this sense a 4-year-old child knows his mother's tongue and so does an adult who has an IQ of 40. Deaf persons are called deficient in language precisely because they do not have readily available a verbal medium in which they can communicate.

While overt language or mastery of language is easily defined, the term language is sometimes used in a sense in which it is difficult to observe its presence or even clearly conceptualize its meaning. Covert or *inner language* is a case in point. It is quite difficult to observe and it is even more difficult

TABLE 1

NONVERBAL COGNITIVE STUDIES WITH THE DEAF

Task and age	Author	Performance of deaf compared to hearing: Not inferior	Inferior
Conceptual-abstract			
Attainment			
10–20	Templin, 1950	Classification	Analogies
5–7	Oléron, 1957	Spatial	Alternation
7–12	Furth, 1961a	Same, Symmetry	Opposite
4–7	Oléron, 1962	Same, Different	
16	Kates, Yudin, & Tiffany, 1962	Strategies	
7–17	Stafford, 1962		Discovery
8–13	Ewing, 1942; Oléron, 1949; Seifert, 1960		1938 Matrices
4–12	Smith, 1952; Farrant, 1964	1947 Matrices	
Transfer			
8–15	Hoefler, 1927; McAndrew, 1948; Oléron, 1951		Shifting
8–12	Rosenstein, 1960	Shifting	
5–7	Oléron, 1957	Transposition	
5, 7, 13	Russell, 1964; Youniss, 1964	Reversal shift	
6, 9	Furth & Youniss, 1964	Reversal association	
6–8	Vincent, 1957		Transfer
Adult	Furth, 1963a	Concept level	Mediated transfer
Adult	Furth, 1964a	Transfer disjunction	
Conceptual-concrete			
Sorting			
6–12	Heider & Heider, 1940	Color sorting	
16, Adult	Kates et al., 1961, 1962	Object sorting	
Knowledge of classes			
8	Vincent, 1959		Classification
8, 12	Rosenstein, 1960	Classification	
8, 16	Furth & Milgram, 1965	Classif., 16 yrs.	Classif., 8 yrs.
6–14	Furth, 1963b	Concept control	
Piaget-type			
5–8	Borelli, 1951	Seriation	
7–17	Oléron & Herren, 1961; Furth, 1964b		Conservation
8	Bradshaw, 1964		Transitivity
Practical intelligence			
6–12	Chuillat & Oléron, 1955		Manipulatory
5–7	Oléron, 1957; Seifert, 1960	Manipulatory	
Memory			
7–12	Blair, 1957	Nonspan	Span
14, Adult	Olsson, 1963	Design span	Digit span
8–11	Doehring, 1960	Spatial	
6–12	Furth, 1961b	Rote, 6–10 yrs.	Rote, 11–12 yrs.
Visual perception			
8–10	Myklebust & Brutten, 1953		Organicity-type
7–10	McKay, 1952; Hayes, 1955; Larr, 1956	Organicity-type	
8, 14	Furth & Mendez, 1963	Gestalt laws	

to define. It can mean—and really should mean nothing else but—silent verbal language. However, at times, the term inner language is used to cover practically the whole domain of cognitive or symbolic behavior and then it is obviously impossible to use it as an explanatory scientific construct.

Terms like *concept* and *symbol* are highly ambiguous and potentially confusing. Language behavior may coincide with conceptual and symbolic behavior, but not necessarily. A person pronouncing nonsense words or words in a language unknown to him engages in verbal but not in symbolic behavior. A child may understand the word "sufficient" as a verbal symbol within a restricted verbal context but have a very inadequate concept of its meaning. On the other hand, one may have an adequate concept of what is expressed by the word "money" yet be unable to define it in verbal terms.

The term intelligence has never been satisfactorily defined so as to find broad acceptance among psychologists, but it is obviously related to conceptual and

symbolic behavior. *Intelligence* is here understood as the ability to perform complex learning tasks. These tasks are commonly concerned with discovering a relevant aspect of the total stimulus situation or with behavior controlled by a rule which may be symbolized and verbalized through a generalized concept.

While it is readily admitted that in everyday life verbal language is usually connected with behavior covered by the above terms, two points should be equally obvious. First, language can be conceptually separated from intelligence; secondly, in cases where language is given an explanatory or indispensible role in intellective development, such a position amounts to a major theoretical assumption. It is here suggested that any theory of cognition which invokes language as an explanatory variable should be tested against the behavior of deaf children and adults.

Conceptual Tasks with Abstract Material

An overview of the majority of studies is presented in Table 1 in the order in which they are discussed in the body of the article. The sections correspond to headings in the text and are there explained. Besides giving the age level of subjects employed in the studies, the last two columns of Table 1 indicate the nature of the task and the relative performance of deaf and hearing subjects. It should be noted that by the term "nonverbal" studies nothing is implied but the minimal use and requirements of verbal behavior in instruction, performance, or as criterion of success.

Nonverbal conceptual tasks can be conveniently classified into those employing relatively meaningless artificial materials, such as geometrical designs, blocks, etc., and those which use meaningful material, familiar from everyday life for which verbalization may be readily available. The stimulus material of the first group is referred to as "abstract" insofar as relevant stimuli are treated as isolated dimensions in distinction from the latter group which treats stimuli as concrete objects. In both these situations the emphasis of the experiment may be on the discovery aspect or on the use a person makes of a principle he is assumed to know. These rough distinctions will allow a more orderly presentation of research by starting with concept discovery studies employing abstract material and proceeding to tasks emphasizing concept shift and transfer.

Concept Attainment

Studies under this heading have as a major task requirement the discovery of a relevant principle (i.e., a concept) according to which success can be achieved. Templin (1950) employed the Brody Nonverbal Abstract Reasoning Test together with other tasks which used a verbal procedure to test hypotheses about the influence of deafness or residential schooling on reasoning in children. The Brody test was divided into classification and analogy subtests. On the classification problem, subjects had to identify one of five figures which did not belong; on the analogy items, they had to select from multiple-choice alternatives the figure which had a relationship to the third figure analogous to that of the second to the first. With chronological age ranging from 10 to 20 years, and with an N of 106 and 56 for the main comparisons of residential and day schools, respectively, hearing and deaf groups were matched on a one-by-one basis. Templin found no differences between hearing and deaf subjects on the classification task, but on the analogy task the deaf

were below the hearing, particularly the residential deaf. The interpretation of this latter result is somewhat shaky because of the uncertainty that the severely deaf understood the instructions. Templin (1950) believed that the results, in general, indicated equality of reasoning ability between deaf and hearing "as long as the general type of reasoning measured is one with which the subjects are familiar."

Oléron of France is one of the few psychologists to show a continuing interest over the years in the cognitive development of deaf children and in theoretical implication for the language-cognition relationship. In 1957 he published a major monograph which examined deaf children's performance on nonverbal types of cognitive tasks which were hypothesized as possibly benefiting from language. In Chapter 4 of this monograph, Oléron (1957) reported his observation with 33 deaf children, aged 4–7, on tasks of what he called temporal and spatial order. The temporal task was a double and a triple alternation problem on which the deaf were about 2 years behind hearing children. On the discovery of a spatial order no difference between deaf and hearing children emerged. From these results, Oléron argued in a post hoc fashion that alternation tasks probably require use of symbols (i.e., language) to a greater degree than do spatial-order tasks although there was no a priori reason to assume that one task required symbols less than the other.

Furth (1961b) studied the attainment of three concepts with the expectation that lack of language would not handicap deaf children in discovering the principles of Sameness or Symmetry but would do so in attaining and using the concept of Opposition. Results supported the expectation: on the first two concepts no consistent difference emerged, while on Opposition, deaf were poorer than hearing children at all age levels, 7 through 12. Deaf children's inferiority on one specific concept was predicated upon the assumption that verbal language by the constant use of opposites gives the hearing children an advantage over the deaf. This seemed an appropriate example of what Vygotsky (1962) called pseudoconcepts, created through linguistic usage. By this is meant behavior, mostly verbal, which gives the impression of mastery of a certain concept, but which under close scrutiny turns out to be mainly determined by frequent linguistic associations. Thus, a child may use the word "sufficient" in many correct contexts but when urged to distinguish "full" from "sufficient" he may reveal his relative ignorance of the mature concept. In the same way it is not likely that the 7-year-old children of this experiment really mastered the concept of opposite, but it is more reasonable to suppose that their test performance was helped by thorough familiarity with their mother tongue.

Consistent with the above results was a study by Oléron (1962) reporting no appreciable differences between deaf and hearing children on learning tasks based on the concepts of Sameness or Difference. Comparing performance of 38 deaf and 38 hearing children, aged 4–7, with stimulus-discrimination objects of shape, object, color, weight, size, or speed, Oléron observed that from age 5 onward a majority of both groups of children succeeded.

Kates, Yudin, and Tiffany (1962) studied attainment of conjunctive concepts by 30 seniors at the Clarke School for the Deaf. The deaf seniors were paired in two ways with hearing controls: on IQ and age or on IQ and scholastic achievement. The materials were cards illustrating various combinations of four attributes grouped in an

orderly or random fashion and six concepts were chosen as focus cards for six different problems. Deaf and hearing subjects were found to proceed by use of similar strategies and no other differences emerged from among a great number of comparisons made except that deaf subjects took more time in making a first choice.

Stafford (1962) paired 29 deaf students, aged 6–17, with hearing controls and presented them with 12 nonverbal discovery problems in increasing order of complexity. Unfortunately the study is not reported in sufficient detail. The summarized results mention superiority of hearing subjects on number of problems solved and trials to criterion per problem solved except that the older deaf required less trials than the hearing successful subjects.

Three investigations in three different countries used Raven's 1938 Progressive Matrices individually on a large number of deaf students: Ewing and Stanton (1943) in England, Oléron (1949, 1950) in France, and Seifert (1960) in Germany. Only Ewing used a hearing control group while the others relied on Raven's tentatively established norms. All three investigators found that only a small percentage of subjects (between 6 and 11%) fell into the two above-average categories (25% expected) while between 15 and 27% were in the lowest subnormal category (5% expected). Ewing explained part of these findings on the basis that schools for the deaf invariably have more subnormal children than normal schools from which defective children are more commonly excluded. While Oléron considered the results as evidence for a developmental retardation in deaf children, Seifert pointed out that after age 12 the deaf seem to make up to a great extent for the developmental retardation shown at earlier ages.

Smith (1952) presented the Colored Progressive Matrices to all younger pupils ($N = 28$, age 4–10) enrolled at a school for the deaf and found that they performed similar to published norms; but the 30 older pupils, tested on the adult version of the test, showed a lower mean score than the norms. Farrant (1964) compared 120 deaf or hard-of-hearing subjects between the ages of 8 and 12 with a hearing control group on a battery of tests which also included the Colored Progressive Matrices. In his factor-analytic treatment of the data, the author observed that the deaf were retarded on all verbal tests and on some tests of figural reasoning but they were not different from the hearing on the Raven test.

Concept Transfer

The ability to classify same objects successively under two or three different aspects has long been regarded as a mark of intelligent, abstract behavior. In particular, rigidity in shifting was considered indicative of aphasic or general cognitive dysfunction. Four studies have been directly concerned with the problem whether nonspeaking deaf people are in this one respect of rigidity alike to aphasic patients. Hoefler (1927) presented the Weigl Sorting Test to 30 deaf school children and observed that only four shifted spontaneously from one sorting category to another. McAndrew (1948) in a restructuring situation employed various degrees of social pressure to have his subjects, ranging in age from 9 to 15 years, change their previous principle of classification. Only 4 out of 24 deaf students changed while all the normals easily succeeded in doing so. Oléron (1951, 1953) worked with a somewhat older age group (mean

age 15) and gave training trials with blocks and transfer trials with drawings. Differences between deaf and hearing were not as marked as with the two former investigators. Yet more hearing subjects shifted spontaneously to a third sorting on the transfer task but the deaf responded easily to suggestion, hence, there were few total failures to shift on their part. The order of shifting, object, color, number, was predominant among the hearing but no one order was characteristic for the deaf. Oléron interpreted the deficiencies in terms of a lack of subordination of perceptive elements to conceptual condition or of less differentiation of possible relevant perceptual aspects. These characteristics were thought to be due to general lack of linguistic habits and not at all related to any aphasiclike rigidity.

Rosenstein (1960) employing the Wisconsin Card Sorting Test with 8–12 year old deaf children found no differences at all between his 60 deaf and 60 hearing control subjects as measured by total errors, number of concepts attained, or perseverative errors. Other investigators cited in this and subsequent sections also used shifting as part of their experimental procedure insofar as any transfer design requires shifting. They likewise failed to observe any particular shifting difficulty in deaf subjects.

In Chapter 3 of his monograph, Oléron (1957) reported on transposition tasks which he gave to deaf children since a number of earlier studies had suggested that language ability may be a decisive determinant in such tasks. Transfer tasks were devised for size alone, size with form changed, weight, and speed. On the first three tasks no differences were observed between the performances of 4–7 year old deaf and hearing children. On the speed task, results were inconsistent with the deaf, apparently poorer in one but better on a second measure. In spite of the small number of subjects, Oléron's major conclusion was that attainment and use of relation is not adversely affected by deafness and its concomitant language deficiency.

In order to fit complex learning processes into a behavioristic model verbal mediation has been postulated as a crucial variable. It was suggested that verbal mediation was related to the performance of reversal as compared to nonreversal learning on the grounds that presence of a mediator facilitated reversal but absence made reversal more difficult than nonreversal shift. Insofar as the mediator is conceptualized as verbal in nature, deaf as compared to hearing children may be handicapped on tasks in which verbal mediation is presumed to operate because of this impoverished linguistic experience. The following two studies tested the reversal-nonreversal paradigm with deaf children. Russell (1964) replicated a reversal study of Kendler and Kendler (1959), and Youniss (1964) employed an extended design of two shift stages. In both studies there were no differences between hearing and deaf children on reversal shift performance. The conclusion to be drawn here from these studies would suggest that mediated behavior is not to be too readily identified with verbal behavior, and that lack of early linguistic experience does not retard such behavior (cf. Youniss & Furth, 1963).

Within a different context, Furth and Youniss (1964) studied paired-associate learning of color and colored objects under reversed and control condition in 6- and 9-year-old children. For example, in the reversed condition, subjects had to learn such associations as "fire truck-white" and "refrigerator-red"; while in the control condition,

"blue" and "brown" were substituted for "white" and "red," respectively. It was observed that hearing but not deaf subjects' performance was significantly worse in the reversal condition. Far from being "perceptually bound" the deaf apparently paid less attention to the color of the objects and hence were not as much hindered when the natural colors were reversed as were the hearing.

Michèle Vincent (1957), another French investigator who has published several psychological articles on deaf children, gave deaf children, aged 6–8 years, the task of discovering successively three different principles of sorting and finally a series of mixed items without correction as measure of transfer of the previously learned principles. The main results showed that deaf children were about 1 year behind hearing children: They performed more poorly on the discovery task of attaining color and size while disregarding form and on the final transfer task. However, they were not considered to exhibit special difficulties in the shifting aspect of the task. Vincent interpreted her results in Piaget's framework and pointed out that language should not be considered the exclusive way of passing from perceptive to intellectual processes even though as a habit it may facilitate this transition.

Doehring (1960a) presented some observations on color-form attitudes of deaf as compared to hearing children. His hearing subjects ranged in age from 4 to adult years and he found, against expectation, that age had no appreciable influence on the proportion of form-preferred choices on a color-form task in which either choice could reasonably be made. At the 8-year level there were 58% form choices for hearing boys and 66% for girls, while only 36% and 53% of form-preferred choices, respectively, were registered by deaf subjects.

Furth (1963a) performed a transfer experiment which had a bearing on linguistic habits and level of concept attainment. College students made four different kinds of shifts after a first concept-attainment task involving a disjunctive concept. It was found that performance on the second task was in direct relation to a possible verbalization of a rule for the first task; yet if the nature of the disjunctive concept had been mastered during the first task, any of the four subsequent shifts would have been equally easy. Only one shift did not benefit from the verbalization and, in fact, the performance of the hearing college students on this shift was poorest. Deaf college students were equal to hearing students on this shift; they were, however, poorer on the other three shifts presumably because the deaf did not benefit from linguistic habits. Where linguistic habits did not help, namely, on the fourth task, the hearing students showed a level of concept attainment not above that of deaf students.

To test the hypothesis that experience may be a sufficient determinant for the development of intellectual capacities and that deaf adults may have made up their possible initial experiential deficiency by simply living an adolescent and adult life, Furth (1964a) devised a series of nonverbal transfer tasks which were aimed at incorporating the following three controls deemed necessary for use with deaf adults: completely nonverbal procedure, initial training to assure understanding of instruction, and performance of some logical operation which differentiates adult persons of above and below average intelligence. Two sorting tasks were designed and were presented as eight conceptual problems. Problems 6 and 7, the critical problems, were shift tasks in which already familiar

elements were sorted into various conjunctive and disjunctive combinations. On such tasks the more intelligent hearing group performed better than the less intelligent group. Apparently, intelligence was related to the ability to benefit on the critical problems from the foregoing training tasks. The comparison of deaf and hearing control groups showed remarkable similarity in terms of trials to criterion, number of errors, or proportion of successes. It should be added that this study for the first time employed deaf adults aged 20–50 years who constituted a sample of typical noncollege working people.

Conceptual Tasks with Concrete Material

There are four sections under this heading. Common to these is the use of stimulus material that is familiar from everyday life and in this sense the stimulus situation can be called concrete. The first two sections deal with free sorting and with knowledge of classes, respectively, the third is devoted to problems taken from Piaget's investigations and finally the last section deals with problems of practical intelligence.

Sorting

Heider and Heider (1940), as some investigators before and after them, considered the possibility that deafness affects cognitive behavior in a way typical of aphasic patients. Interest in this question is understandable and is bound to come up repeatedly because of the exterior similarity of speech and language defect in deafness and aphasia and because of the presence of the poorly diagnosed multiple-handicapped child in schools for the deaf. Heider and Heider required deaf and hearing children at various age levels to select from many blocks of different hues those that were similar to certain standard colors. On such a task, aphasic subjects typically select only an exact match and reject similar hues as being different. The deaf children in this experiment performed in an opposite fashion, that is, they selected a wider band than hearing controls, a result which is more related to a younger chronological age. However, differences in number of instances chosen, although fairly consistent over all age levels, were small and permitted the assertion that the performance of the deaf children was quite similar to hearing subjects and quite incongruent with aphasic behavior.

Kates, Kates, Michael, and Walsh (1961) and Kates, Kates, and Michael (1962) employed the Goldstein-Scheerer Sorting Test, a task which was also originally designed to probe into cognitive processes affected by aphasic conditions. The authors, as before (Kates, Yudin, & Tiffany, 1962), worked with a matched sample of deaf seniors and two control groups and an additional group of adult alumni from the Clarke School for the Deaf. While the authors' main concern was a comparison of nonverbal and verbal classificatory behavior and some expected differences were found in verbal performance, no differences were observed on nonverbal performance. In this respect the deaf categorized familiar objects at a level comparable to their peers under free and more structured sorting conditions and they manifested as much flexibility and acceptable range of choices. Generalization from these results is only limited by the fact that the deaf subjects were a select group and uniformly well above average ability and linguistic achievement as compared with the average deaf person.

Knowledge of Classes

These studies were nonverbal versions of the familiar Similarity test of the type: "In which way are . . . and . . .

alike?" Vincent (1959) obtained norms for young children on a task in which they had to group familiar objects and afterwards were required to verbalize the reason for their choice. When she observed the grouping behavior of 8-year-old deaf children she found their performance on the level of 6-year-old hearing children. Rosenstein (1960) employed a paired-associate procedure of three lists of five pictures each, illustrating five different concepts. Attainment of concepts was inferred by improved performance on Lists 2 and 3. Deaf and hearing subjects showed similar improvement and the author considered these together with earlier mentioned results as evidence that deaf children do not consistently exhibit conceptual restriction or reduced ability to think abstractly. Furth and Milgram (1965) reported on a picture similarity task with 8- and 16-year-old deaf subjects in comparison with normal and in addition subnormal subjects who also verbalized to it and were given the task in verbal form. The nonverbal task of interest to this discussion had 20 different problems. On each problem a subject had to select three out of seven pictures presented that shared a common attribute. The younger deaf children were found to make more errors than their hearing 8-year-old peers with their performance being comparable to that of 6-year-old children while the deaf adolescents succeeded as well as their counterparts. The authors, relying on evidence from the results with the subnormal group, suggested as probable source for their deficient classificatory behavior not lack of language but rather the experiential restriction of deaf children's early life resulting from the lack of language. Results with the older group were cited as additional demonstration that no lasting conceptual deficit was present in deaf children.

In order to observe classificatory behavior under the control of one concept, Furth (1963b) collected normative data on a pictorial choice task of which the major source of difficulty was the consistent selection of two pictures illustrating the part-whole concept in spite of the presence of other reasonable choice possibilities. The term "verbal control" has often been employed in connection with increased cognitive control of the growing child (cf. Luria, 1957; Mussen, 1963). If a readily available language influenced controlled behavior, deaf more frequently than hearing children should make more incorrect, perceptually-bound choices. Comparison of the performance of deaf and hearing subjects over all age levels revealed, however, a close correspondence between their scores. Moreover, results indicated that over an age span of from 6 to 14 years intelligence contributed significantly to a better performance of hearing subjects on the part-whole task, age contributed to a smaller degree, and success in verbalizing the concept hardly at all. Since presence versus absence of the verbal label on this and other conceptual tasks was observed to be irrelevant to success, the author suggested "conceptual control" as a more appropriate and more inclusive term for a performance which is under the control of an internal principle rather than verbal control.

Piaget-Type Tasks

Piaget's theoretical model of cognitive development and his basic biological position on intelligence as being rooted in overt action would suggest that the growth of intelligent, logical operation is not dependent on but rather reflected in language behavior. In line with this reasoning, it could be predicted that deaf would not differ from hearing children with respect to the age of which logical operations emerge. Within such a theoretical framework, Michèle

Borelli Vincent (see Borelli, 1951) tested children of between 5 and 8 years in half-year steps on their ability to place first manikins and then sticks along a size continuum and finally to place the matching elements of both series in correspondence under various conditions. Success on such tasks was considered by Piaget as an indication that the child is beginning to use the number concept in a logical manner. Vincent observed only a slight retardation of about 6 months in her deaf subjects, as compared to hearing subjects. She concluded that these deaf children exhibited no fundamental difference in logical elementary capacities and that the beginnings of logical operations are largely independent of language. She speculated further that probably with operation at a more formal level of logical thought these conclusions may not hold.

Deaf children's performance on some of Piaget's familiar conservation problems was tested by Oléron and Herren (1961). Rightly arguing that deaf children were handicapped if language was part of the experimental procedure, the authors devised a series of pictures which were learned by their subjects, designed to serve as equivalents of the verbal responses "heavier," "same," "lighter" for the conservation of weight and "more," "less" for the conservation of volume. After training with these pictures, the age at which 50% of hearing succeeded on the weight problem was 8.5 and with the quantity of liquid problem was 10.5. Corresponding age levels for deaf subjects were about 6 years later. While Oléron admitted that the retardation of 6 years need not be taken in an absolute sense he believed that Piaget's theory does not emphasize sufficiently the role of language in the emergence of logical behavior, particularly in the subordination of perceptual to conceptual conditions.

As the possibility existed that the use of pictorial symbols introduced an extraneous difficulty for deaf subjects which was mainly responsible for Oléron's results, Furth (1964b) replicated the conservation experiment. He modified the pretask procedure by training his subjects first to express a judgment of equal weight by a horizontal hand movement and indicate judgment of heavier by moving downward the hand holding the heavier object. Two clay balls were used, transformed before their eyes, and put back into their hands. The lowering of the hand on the critical trials indicated a wrong judgment since the object was merely the transformation of a former object. After observing responses of 8-year-old deaf children it was discovered that their final performance was not like 8-year-old hearing children in Grade 3 but more like 6.5-year-old children in Grade 1. A qualitative investigation of failures, however, revealed that many more 8-year-old deaf children made hesitant and inconsistent responses than 6-year-old hearing children. This was interpreted as a possible clue that the older deaf children felt uncomfortable about the response and were really closer to the correct solution than a mere summary of failures or successes would indicate. Evidence from other sources (Goodnow, 1962; Lovell & Ogilvie, 1961) is cited which corroborates the view that kind of experience with the physical world rather than language or formal training determines in part the age at which children pass from a perceptual to a logical judgment on many Piaget-type experiments.

Bradshaw (1964), in a study of inferential reasoning of size relation, used nonverbal methods with hearing and deaf children, aged 6 and 8 years. His procedures were divided into preoperational and two concrete-operational tasks. The results demonstrated that the older hearing children differed from

the younger on the latter but not the former tasks, and that the 8-year-old deaf children performed more like 6-year-old children on the sequential but were equal to their peers on the simultaneous operational task. The author pointed out that the average inferiority of the deaf group may have been associated with their lower socioeconomic status.

Practical Intelligence

Chuillat and Oléron (1955) reported the results of a series of six experiments based on a standardized test of practical intelligence by Rey. They observed that tasks which were easy to solve for hearing children, aged 6–7, were difficult or insolvable even with help from the experimenter for the deaf school children in an age range from 5 to 12. Such a striking inferiority in practical intelligence was not replicated in a second study (Oléron, 1957, Ch. 2) in which deaf children, aged 5–7, solved a manipulatory problem quantitatively and qualitatively very much like hearing children of their age. While Oléron in explaining the first results had recourse to factors such as understanding of instructions, personal initiative and language habits in structuring a task, the present finding seemed to him to contradict a rigid theory which linked language and practical intelligence.

Seifert (1958) observed that a sample of 80 deaf children were as skilled as their hearing peers in manual dexterity in cutting out figures and bending wires according to a given copy.

Tasks of Memory

Blair (1957) compared the performance of two groups of 53 deaf and hearing children, aged 7–12, on visual tasks of immediate memory: Knox Cube, memory for design, object location, and four different span tasks. Only overall mean scores were presented with age levels collapsed. While the two groups did not differ on object location, deaf subjects were superior on Knox Cube and memory for design but consistently inferior to hearing subjects on the span tasks. The author suggested that such a performance pattern may be characteristic for the deaf child's visual memory as a compensatory effect for his auditory deficiency. Doehring (1960b) failed to observe any notable difference between hearing and deaf children, aged 8–11 on a memory task for spatial location under various conditions of delay, interference, and exposure times.

Olsson (1963) tested adolescent and adult subjects on visual span tasks and observed the following results: on visual nonsense figures, deaf and hearing subjects, in general, performed alike on simultaneous and successive presentation of nonsense figures; on visual digits, however, deaf subjects were poorer than hearing subjects at both age levels. Moreover, when recall of figures of low and high associations was compared, deaf subjects against expectation improved as much as hearing subjects on high association figures. This similarity between deaf and hearing subjects with figures of high and low association value would seem to demonstrate that a deficient memory performance on the part of the deaf cannot simply be due to their verbal deficiency. Otherwise one would have expected that a high verbal association value would be of less facilitation to deaf than to hearing subjects. The differences in the performance on digits were interpreted as manifesting the particular effects of continual practice of hearing individuals to a much greater degree than deaf persons in recalling digits (e.g., telephone).

Naffin (1959) mentioned pertinent research published in German journals and books for teachers of the deaf. This research, Naffin noted, demonstrated a marked superiority of the deaf child's visual memory in certain specific situa-

tions which were characterized by a perceptual poverty or simplicity.

Furth (1961b) presented a paired-associates task with four visual color stimuli and two motor responses. A comparison of total errors yielded no differences on this rote performance between hearing and deaf children aged 7–10, but superiority of hearing subjects aged 11 and 12.

Tasks of Visual Perception

Myklebust and Brutten (1953) used a series of visual perception problems: marble board, figure-ground, pattern reproduction, and perseveration. These tasks like others mentioned earlier, originally designed to observe behavioral disturbances in brain damage, were employed to investigate the possibility of organicity-like disturbances in deaf children. The authors reported that their deaf subjects, aged 8–10, performed significantly different from their hearing controls on a number of tasks: they made more errors in reproducing marble patterns present before their eyes, in two out of five problems they required more exposure time to succeed in reproducing patterns of dots, and a greater number of deaf subjects gave background rather than foreground responses in one out of five items in the object series and three out of five items in the design series. No statistical differences between deaf and hearing groups emerged on reproduction of line patterns and on perseverative after-effects with an ambiguous figure. The authors concluded that the results supported their initial hypothesis of an organismic disturbance due to sensory deprivation in deaf children and that their perceptual inferiority is related to an impairment in abstract functioning.

In view of these striking differences and sweeping implications it seemed important to replicate the above study. Larr (1956) performed this task with two independent groups of 25 deaf students each from the lower grades of two different schools for the deaf. The deaf groups were paired with a normal and subnormal control group. Larr's results showed that the deaf subjects performed equally well on marble board reproduction. On the figure-ground task a greater proportion of deaf subjects gave background references than did either control groups. Such findings seem to weaken the evidence for a general perceptual disturbance in deaf children, particularly since two more investigations (Hayes, 1955; McKay, 1952) also reported results not in agreement with those of Myklebust and Brutten.

Furth and Mendez (1963) studied visual perception of figures which illustrated some of the major principles of gestalt. Deaf and hearing subjects at two age levels (9 and 16 years) performed, in general, alike except that younger but not older subjects were inferior to hearing subjects on two problems which called primarily for a discriminatory as opposed to a global approach.

Deaf Person's Performance on Cognitive Tasks

The above summarized studies of deaf people's nonverbal cognitive performance do not provide a unified picture; in each topic discussed conflicting results were found. Some of the factors which may be responsible for this lack of agreement are rather obvious. Nonverbal procedures which are not uniformly understood by subjects—noncognitive motivational or emotional characteristics, feelings of insecurity in a problem situation with no verbal explanation, and other factors—may interfere with good cooperation. There is finally the ever present personal bias of the investigator who either expects or does not expect to find differences in performance between deaf and hearing subjects and who according to this ex-

pectation selects the problems and methodological procedures. However, if these things are taken for granted as part of any human endeavor, there remains a considerable number of studies in which deaf subjects performed quite similarly to hearing subjects. These empirical results would be evidence for the assertion that the cognitive development and functioning of deaf people—outside of the realm of language—is not different from hearing people.

One cannot ignore, however, these studies in which deaf subjects performed below the standard of hearing peers. Where such reported differences are reliable, the investigator is left with the task of explaining these deficiencies. In some cases, differences were expected and directly related to lack of linguistic experience (Furth, 1961b) and in some others this relationship was only observed post facto (Furth, 1963a, Olsson, 1963).

In other studies the relationship between lack of language and present inferior performance of deaf subjects was more a matter of speculation than of observation (e.g., Furth, 1964b; Oléron, 1951, 1957; Vincent, 1959). The danger of coming up with untestable hypotheses and of circular reasoning is rather obvious in the latter situation. Concerning the retarded performance of deaf children, even on some nonverbal tasks, Ervin and Miller (1963) cautioned wisely that other than linguistic factors, such as a great proportion of psychologically or mentally abnormal or brain-damaged children in a deaf sample, may affect the results. They imply that one should be wary about inferring some direct or indirect effect of language when other variables could just as easily or better explain an average deficiency with a deaf sample.

In line with this reasoning the present author has on different occasions (Furth, 1961a, 1964b; Furth & Milgram, 1965) proposed that lack of language experience may have a retarding effect on the individual child indirectly via lack of sufficient cognitive stimulation and motivation. This possibility is more than mere speculation since it is common knowledge among workers in the field that deaf children rarely enjoy cognitive stimulation and social-emotional acceptance equal to hearing children (cf. Levine, 1960). Deaf children are, in fact, usually neglected children in some respects. After making this concession to language, it is suggested that this unfavorable set of environmental circumstances is remediable. A possible change in overall educational attitude could encourage the early use of nonverbal cognitive stimulation and learning and thus could largely compensate for deficient linguistic skill. Such a hypothesis can be subjected to experimental testing, although the difficulty of executing this design is not underestimated.

An explanation of experiential deficiency would thus seem to be a parsimonious one in understanding the results of a number of cognitive studies, particularly those in the sections on Knowledge of Classes and Piaget-Type Tasks which reported a lag of 1 or 2 years between deaf and hearing young children. The fact that with increasing age less or no difference was observed by several investigators (Furth & Mendez, 1963; Furth & Milgram, 1965; Seifert, 1960), or that with improved controls previously reported differences were no longer manifest (Rosenstein, 1960; Larr, 1956), or that deaf adults who were not skilled in language performed as well as hearing adults on classification tasks involving rather difficult conjunctive and disjunctive concept shifts (Furth, 1964a), would appear to lend additional support to the suggested explanation.

With the studies in which deaf subjects performed at a level of hearing peers, the above mentioned argument

of Ervin and Miller may be turned around and one would seem to have empirical reasons for stating that, at least, linguistic skill or deficiency is not an important determinant of the particular task at hand. While such studies may not be particularly enlightening as to deaf people's intellectual functioning, they may be helpful in clarifying the status of some hypothetical constructs related to language, such as verbal mediation, symbol and concept formation, concrete and abstract behavior, etc. Rosenstein (1961), discussing the literature on cognitive research in deaf children, has highlighted some of the confusion which stems from lack of a clear terminology. Some of these points have been briefly touched upon during the discussion of specific research, for example, on transposition (Oléron, 1957), reversal (Youniss, 1964), conceptual control (Furth, 1963b), logical operations (Borelli, 1951), verbal association (Olsson, 1963).

Implication for the Relationship of Language and Intelligence

By generalizing the results of the studies summarized above and applying them to a theoretical position on the influence of language on intellective development, the following is suggested: (*a*) Language does not influence intellectual deveolpment in any direct, general, or decisive way. (*b*) The influence of language may be indirect or specific and may accelerate intellectual development: by providing the opportunity for additional experience through giving information and exchange of ideas and by furnishing ready symbols (words) and linguistic habits in specific situations.

From this position it should follow that persons, deficient in linguistic experience or skill (*a*) are not permanently or generally retarded in intellectual ability, but (*b*) may be temporarily retarded during their developmental phase because of lack of sufficient general experience and (*c*) they may be retarded on certain specific tasks in which available word symbols or linguistic habits facilitate solution.

When the question is raised: "How do deaf persons think, if not in language?" the person asking the question may understand by language any kind of symbolic or communicative behavior and thinking may include for him any internal activity related to intelligent behavior. However, language and intelligence are here taken in the specific sense in which they were first defined. Since there is general agreement that most of the learning studies reviewed could not be solved in the absence of symbolic activity, successful performance on these tasks by deaf persons implies an efficient functioning of a symbolic system other than verbal. If the above question is then modified to read "What kind of symbols do deaf people use?" the author is doubtful that this kind of question lends itself to an empirical answer. A search for explanation of conceptual behavior in terms of symbols may lead closer than is realized to the impasse of the early introspectionists who were looking for the elements which make up thinking.

In general, it should be understood that the social and cultural aspects of language have been taken for granted throughout this discussion and no speculations about the role of language in creating and transmitting a culture are made.

It is quite possible, for instance, that a deaf person may act in a way which seems unintelligent to a hearing person, but the reason for his action may be in the deaf person's different kind of experience and information on which he premises his action. By the same token, it may well be that only by extensive perusal of verbal material a person can

be motivated and become directly interested in some specific intellective pursuits. Finally, it would be absurd to deny that hearing people commonly employ language to express and communicate intelligent acts overtly and often use some form of language internally when searching for an intelligent understanding of a given situation. In all these cases, however, language seems to play a subsidiary role to intelligence proper, a role which the reviewed studies indicate could in some instances be taken over by other symbol-like structures.

It may be necessary to clarify more precisely what is implied by the term "the indirect influence of language" since Oléron (1957) employed a similar phrase with a somewhat different meaning. For Oléron, language is directly supposed to further and train intellectual experience, for example, by saying a "brown bookcase" the child is led to perform the intellectual operation of abstracting the concept of "brownness" from the concept of "bookcase" and by constant use of such linguistic habits which separate a noun from an adjective he develops a corresponding intellectual habit. According to Oléron, language makes an essential contribution by helping the person to transcend the perceptual order and by giving habitual training in intellective habits and symbolic techniques, even if it is not the main determinant of intellective development. Although the distinction may appear subtle, in our discussion the notion of an indirect linguistic influence is used in a different sense.

Our position refers to the human person in his interaction with the environment—which is called experience—as being responsible for intellective development. A hearing child through language may simply have more opportunity to interact with or meet the environment. Language thus affords opportunity for more experience, but it is not considered a primary or necessary factor in developing intellective habits. In other words, an intellective habit is not thought to presuppose or have at its base a linguistic habit. If it were otherwise, would it not follow that people, deprived of language during the formative years, would remain permanently crippled in intellectual development? Would one not also expect that usage and comprehension of language structure is closely associated with intelligence?

Obviously more refined and sophisticated research with deaf people is desirable to extend the empirical evidence available thus far. Even more important is the quest for other than verbal models to help the psychological investigator understand intellective functioning. A close study of children's nonverbal cognitive development would be most desirable. The reason for stressing the attribute "nonverbal" is simply that if we direct our attention to manifestations of "verbal intelligence" we already preempt the question about the relationship of language and intelligence. Even Piaget (1962) in a recent review of his position admits "that the root of logical operation lies deeper than the linguistic connections, and that my early study of thinking was centered too much on its linguistic aspect [p. 5]."

Where the above theory mentions persons deficient in linguistic experience or skill, other than deaf persons are clearly in mind. Mentally retarded, culturally deprived, speech impaired, and some emotionally disturbed people may all show the specific indirect results of language deficiency. Wepman (in press) and Eisenson (1963) explicitly pointed out that aphasia should be regarded as a linguistic and not a general cognitive handicap. Lenneberg (1962)

presented a case study of an 8-year-old child who had never talked but, in general, understood language and performed on psychometric tests according to his age level, a rather obvious and not too rare instance which demonstrated at least that active speech is no prerequisite for mastery of language and normal cognitive development. Milgram and Furth (1963), replicating the above-cited results with Same, Symmetry, and Opposition (Furth, 1961b) on mentally retarded, suggested that retarded persons can be considered as linguistically deficient over and above their general mental age level.

As a final and complementary body of evidence for the general position of the relative independence of language and intellective development correlations between intelligence and language measures from different authors (Goda & Griffith, 1962; Harrell, 1957; Templin, 1957; Winitz, 1959) were examined. Based on combined results of 746 persons from 5 to 21 years of age the r for IQ and Sentence Complexity was a weak .16 and based on $N = 426$ the r for IQ and Sentence Length was an equally unimpressive .19. Similarly Burt (1949), reviewing the results of several studies, pointed out that there seems to be general agreement among factor analysts that verbal ability is independent of so-called "higher mental processes," such as generalization, abstraction, judgment, and reasoning.

Three additional points may be added, one that hearing children of quite low mentality learn to master the basic structure of their mother tongue at a functional level. Secondly, there is the observation that the learning of foreign languages is not substantially related to measures of intelligence (Carroll, 1962).

Finally, there is increasing evidence that training in verbalizing particular task variables does not, as such, appreciably improve a conceptual performance (Beilin & Franklin, 1962; Wohlwill & Lowe, 1962). In those cases where verbalization did seem to help, inspection of the procedure usually shows that verbalization did more than merely provide words. For example, in Gagné and Brown (1961) all groups were supplied with appropriate verbal terms, but the group which emerged as best learners was guided in using the terms in varying contexts. Thus, this group differed from others not so much in verbal as in conceptual training. Perhaps this is what Gagné implied when he stressed the importance of what has been learned in a situation in contrast to how it has been learned, that is, whether verbally or nonverbally.

In summary, then, the reported investigations seem to emphasize as legitimate the distinction between intellective and verbal skills. The ability for intellective behavior is seen as largely independent of language and mainly subject to the general experience of living. Various sources of empirical evidence confirm the theoretical position that just as language learning is not closely related to intellectual endowment so intellective performance is not directly dependent on language.

REFERENCES

Beilin, H., & Franklin, Irene C. Logical operations in area and length measurement: Age and training effects. *Child Develpm.*, 1962, **33**, 607–618.

Binet, A., & Simon, T. Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *Année psychol.*, 1905, **11**, 191–244.

Blair, F. X. A study of the visual memory of deaf and hearing children. *Amer. Ann. Deaf*, 1957, **102**, 254–263.

Borelli, Michèle. La naissance des operations logiques chez le sourd-muet. *Enfance*, 1951, **4**, 222–238.

Bradshaw, D. H. A study of inferred size relations using nonverbal methods. Unpublished doctoral dissertation, Catholic University of America, 1964.

Burt, C. The structure of the mind: A review of the results of factor analysis. *Brit. J. educ. Psychol.*, 1949, **19**, 176–199.

Carroll, J. B. *The prediction of success in intensive foreign language training.* Pittsburgh: Univer. Pittsburgh Press, 1962.

Chulliat, R., & Oléron, P. Sur le développement de l'intelligence practique chez les enfants sourds. *Enfance,* 1955, **8**, 281–306.

Doehring, D. G. Color-form attitudes of deaf children. *J. speech hear. Res.*, 1960, **3**, 242–248. (a)

Doehring, D. G. Visual spatial memory in aphasic children. *J. speech hear. Res.*, 1960, **3**, 138–149. (b)

Eisenson, J. Aphasic language modifications as a disruption of cultural verbal habits. *ASHA*, 1963, **5**, 503–506.

Ervin, Susan, & Miller, W. H. Language development. *Yearb. Nat. Soc. Stud. Educ.*, 1963, **62**, 108–143.

Ewing, A. W. G., & Stanton, D. A. G. A study of children with defective hearing. *Teacher Deaf,* 1943, **41**, 56–59.

Farrant, R. H. The intellective abilities of deaf and hearing children compared by factor analyses. *Amer. Ann. Deaf,* 1964, **109**, 306–325.

Furth, H. G. Influence of language on the development of concept formation in deaf children. *J. abnorm. soc. Psychol.*, 1961, **63**, 386–389. (a)

Furth, H. G. Visual paired-associates task with deaf and hearing children. *J. speech hear. Res.*, 1961, **4**, 172–177. (b)

Furth, H. G. Classification transfer with disjunctive concepts as a function of verbal training and set. *J. Psychol.*, 1963, **55**, 477–485. (a)

Furth, H. G. Conceptual discovery and control on a pictorial part-whole task as a function of age, intelligence, and language. *J. educ. Psychol.*, 1963, **54**, 191–196. (b)

Furth, H. G. Conceptual performance in deaf adults. *J. abnorm. soc. Psychol.*, 1964, **68**, in press. (a)

Furth, H. G. Conservation of weight in deaf and hearing children. *Child Develpm.*, 1964, **34**, 143–150. (b)

Furth, H. G., & Mendez, R. A. The influence of language and age on gestalt laws of perception. *Amer. J. Psychol.*, 1963, **76**, 74–81.

Furth, H. G., & Milgram, N. A. The influence of language on classification: A theoretical model applied to normal, retarded, and deaf children. *Genet. Psychol. Monogr.*, 1965, in press.

Furth, H. G., & Youniss, J. Color-object paired-associates in deaf and hearing children with and without response competition. *J. consult. Psychol.*, 1964, **28**, 224–227.

Gagné, R. M., & Brown, L. T. Some factors in the programming of conceptual learning. *J. exp. Psychol.*, 1961, **62**, 313–321.

Goda, S., & Griffith, B. C. Spoken language of adolescent retardates and its relation to intelligence, age, and anxiety. *Child Develpm.*, 1962, **33**, 489–498.

Goodnow, Jacqueline J. A test of milieu effects with some of Piaget's tasks. *Psychol. Monogr.*, 1962, **76**(36, Whole No. 555).

Harrell, L. E. A comparison of the development of oral and written language in school-age children. *Child Develpm. Monogr.*, 1957, No. 22.

Hayes, G. A study of the visual perception of orally trained deaf children. Unpublished master's thesis, University of Massachusetts, 1955.

Heider, F. K., & Heider, Grace M. A comparison of color sorting behavior of deaf and hearing children. *Psychol. Monogr.*, 1940, **52**(1, Whole No. 232), Ch. 2.

Hoefler, R. Ueber die Bedeutung der Abstraktion fuer die geistige Entwicklung des taubstummen Kindes. *Z. Kinderforsch,* 1927, **33**, 414–444.

James, W. *Principles of psychology.* New York: Holt, 1890.

Kates, S. L., Kates, W. W., & Michael, J. Cognitive processes in deaf and hearing adolescents and adults. *Psychol. Monogr.*, 1962, **76**(32, Whole No. 551).

Kates, S. L., Kates, W. W., Michael, J., & Walsh, T. M. Categorization and related verbalizations in deaf and hearing adolescents. *J. educ. Psychol.*, 1961, **52**, 188–194.

Kates, S. L., Yudin, L., & Tiffany, R. K. Concept attainment by deaf and hearing adolescents. *J. educ. Psychol.*, 1962, **53**, 119–126.

Kendler, Tracy S., & Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58**, 56–60.

Larr, A. L. Perceptual and conceptual ability of residential school deaf children. *Except. Children,* 1956, **23**, 63–66.

Lenneberg, E. H. Understanding language without ability to speak: A case report. *J. abnorm. soc. Psychol.*, 1962, **65**, 419–425.

Levine, Edna S. *The psychology of deafness.* New York: Columbia Univer. Press, 1960.

Lovell, K., & Ogilvie, E. A. A study of the conservation of weight in the junior school child. *Brit. J. educ. Psychol.*, 1961, **31**, 138–144.

Luria, A. R. The role of language in the formation of temporary connections. In B. Simon (Ed.), *Psychology in the Soviet*

Union. Stanford: Stanford Univer. Press, 1957.

McAndrew, H. Rigidity and isolation: A study of the deaf and the blind. *J. abnorm. soc. Psychol.,* 1948, **43**, 476–494.

McKay, E. B. An exploratory study of the psychological effects of severe hearing impairment. Unpublished doctoral dissertation, Syracuse University, 1952.

Milgram, N. A., & Furth, H. G. The influence of language on concept attainment in educable retarded children. *Amer. J. ment. Defic.,* 1963, **67**, 733–739.

Mussen, P. H. *The psychological development of the child*. Englewood, N. J.: Prentice-Hall, 1963.

Myklebust, H., & Brutten, M. A study of the visual perception of deaf children. *Acta oto-laryngol., Stockholm,* 1953, Suppl. No. 105.

Naffin, P. Die psychologischen Voraussetzungen der Erziehung des taubstummen Kindes. In H. Hetzer (Ed.), *Paedagogische Psychologie.* Gottingen, Germany: Hogrefe, 1959.

Oléron, P. Etude sur les capacites intellectuelles des sourds-muets. *Année psychol.,* 1949, **47–48**, 126–155.

Oléron, P. A study of the intelligence of the deaf. *Amer. Ann. Deaf,* 1950, **95**, 179–195.

Oléron, P. Pensee conceptuelle et langage: Performances comparées de sourds-muets et d'entendants dans des epreuves de classement multiple. *Année psychol.,* 1951, **51**, 89–120.

Oléron, P. Conceptual thinking of the deaf. *Amer. Ann. Deaf,* 1953, **98**, 304–310.

Oléron, P. *Recherches sur le développement mental des sourds-muets*. Paris, France: Centre National de la Recherche Scientifique, 1957.

Oléron, P. Le développement des réponses a le relation identité-dissemblance: Ses rapports avec le langage. *Psychol. Franç.,* 1962, **7**, 4–16.

Oléron, P., & Herren, H. L'acquisition des conservations et le langage: Etude comparative sur des enfants sourds et entendants. *Enfance,* 1961, **14**, 203–219.

Olsson, J. E. The influence of language experience on visual memory span. Unpublished master's thesis, Catholic University of America, 1963.

Piaget, J. *Comments on Vygotsky*. Cambridge: Massachusetts Institute of Technology Press, 1962.

Rosenstein, J. Cognitive abilities of deaf children. *J. speech hear. Res.,* 1960, **3**, 108–119.

Rosenstein, J. Perception, cognition and language in deaf children. *Except. Children,* 1961, **27**, 276–284.

Russell, J. Reversal and nonreversal shift in deaf and hearing kindergarten children. Unpublished master's thesis, Catholic University of America, 1964.

Seifert, K. H. Untersuchungen zur Frage der Kompensation auf dem Gebiet des Handgeschicks. *Z. exp. angew. Psychol.,* 1958, **5**, 465–490.

Seifert, K. H. Erfahrungen mit dem Progressive Matrices Test bei taubstummen Kindern. *Z. exp. angew. Psychol.,* 1960, **7**, 255–290.

Smith, D. I. A survey of the intelligence and attainments of a group of deaf children. *Brit. J. Psychol.,* 1952, **22**, 71–72.

Stafford, K. Problem-solving ability of deaf and hearing children. *J. speech hear. Res.,* 1962, **5**, 169–172.

Templin, Mildred C. *The development of reasoning in children with normal and defective hearing*. Minneapolis: Univer. Minnesota Press, 1950.

Templin, Mildred C. *Certain language skills in children*. Minneapolis: Univer. Minnesota Press, 1957.

Vincent, Michèle. Sur le role du langage a un niveau elementaire de pensée abstraite. *Enfance,* 1957, **10**, 443–464.

Vincent, Michèle. Les classifications d'objects et leur formulation verbale chez l'enfant. *Psychol. Franç.,* 1959, **4**, 190–204.

Vygotsky, L. S. *Thought and language*. Cambridge: Massachusetts Institute of Technology Press, 1962.

Wepman, J. M. Five aphasias: A concept of aphasia as a regressive psycholinguistic phenomenon. *Proc. Ass. Res. Nerv. Ment. Dis.,* **42**, in press.

Winitz, H. Relationships between language and nonlanguage measures of kindergarten children. *J. speech hear. Res.,* 1959, **2**, 387–391.

Wohlwill, J. T., & Lowe, R. C. Experimental analysis of the development of the conservation of numbers. *Child Develpm.,* 1962, **33**, 153–167.

Youniss, J. Concept transfer as a function of shifts, age, and deafness. *Child Develpm.,* 1964, **34**, in press.

Youniss, J., & Furth, H. G. Reaction to a placebo: The mediational deficiency hypothesis. *Psychol. Bull.,* 1963, **60**, 499–502.

(Received January 2, 1964)

Psychological Bulletin
1964, Vol. 62, No. 3, 165–179

STIMULUS CONTROL FOR BIRD ORIENTATION[1]

MERLE E. MEYER
University of Washington

During the last 2 decades there have developed 4 experimental lines of approach toward the understanding of the stimulus control of bird orientation and navigation: sun-compass-orientation, star-navigation, sun-navigation, and what might be called nonvisual hypotheses. Each position has yielded important evidence which has in turn led to sharper definitions of problems and more rigorous appraisal of theory. The intention of this paper is to review the research findings and conclusions concerning each of the positions.

The stimulus control of navigation and migration by birds has constituted one of the most challenging problems for the student of animal behavior. This general topic has always aroused the imaginative and the curious, but it has only been during the past 20 years that impetus has been given to the development of new explanations and new methods of investigation. During this time four lines of approach have emerged. The sun-compass-orientation hypothesis set forth by Kramer and his co-workers has established that diurnal migrating birds maintain their direction by orientation with respect to the sun. With Sauer's work and the development of the star-navigation hypothesis, there is evidence that nocturnal migrating birds orient and navigate by means of star patterns. Matthews' sun-navigation hypothesis has been put forward to account for the evidence that birds can use observations of the sun for true bicoordinate navigation, and lastly the nonvisual hypothesis, orientation and navigation by sensitivities to magnetic and "Coriolis" forces, has been proposed. This review concerns itself with the advances and conclusions made by each position.

Sun-Compass-Orientation Hypothesis

Virtually nothing was known of the mechanisms which organisms use for orientation prior to 1949. Since that time it has been shown that several species of birds, arthropods, fish, and reptiles possess the ability to find compass directions with the help of the sun and to keep that direction despite the "movement" of the sun during the day.

The sun-azimuth orientation consists in the determination of the compass directions by the sun position regardless of the time of day; hence, it requires mainly the ability to correct for the sun's daily azimuth displacement. Generally it has been assumed that only the azimuth of the sun's position is used as the basis for orientation. The sun's azimuth, that is the projection of the sun's position onto the horizon, however, changes during the day with varying speed. In order to maintain one's compass direction accurately, the correction for the azimuth movement has to vary with the same rhythm. The azimuth movement varies, not only during the course of the day, but also with the seasons and with various latitudes. An

[1] This paper was written while the author was on a National Science Foundation Faculty Fellowship and on leave from Whitman College. The writer wishes to express his thanks to Geo. P. Horton, Paul E. Fields, Donald M. Baer, Bert Forrin, Robert Lockard, Gordon H. Orians, and Frank Richardson, all of whom read the paper and made valuable comments.

organism which uses only the sun's azimuth for keeping a compass direction has, therefore, to compensate for different azimuth movements in equal time intervals at different times of the day, at different seasons, and at different latitudes if it migrates.

It has been suggested by Kramer (1961) and by Wallraff (1960a, 1960b) that two kinds of orientation occur in nature, one-direction orientation and goal orientation. In the former case, no matter where the animal is, it tries to go in a given compass direction; in the latter case, it tries to go towards a "home" goal. According to the hypothesis, the direction to be taken is in many cases unlearned, while the location of a goal area, on the other hand, must be learned or imprinted through experience. Navigation is the ability to recognize displacement in unfamiliar surroundings and to choose the correct homing direction.

Kramer (1950a, 1950c) found that his captive starlings *Sturnus vulgaris* which lived in an outdoor aviary got restless at migration time and would cling to the side of the cage or make short sallies in the direction they would take if free to migrate. The starlings' visual field was restricted to that of the sky above by a round aviary which excluded the sight of the horizon. Kramer noted that the diurnal fluttering remained oriented in the species-specific migration direction only when the sun or the sky near it was unobscured by cloud cover. This directional tendency was manifest even in a round cage of a diameter of about *2* feet. Only the overhead sky and the features of the apparatus remained as visual cues and only the former provided constant directional information, since the apparatus was regularly rotated. At the same time, it was clear that the birds were able to correct for the sun's movement during the day, for they showed the ability to take time into account by holding the same direction in spite of passing time. However, individual starlings differed in appraising the hourly angle of displacement in spite of their being constantly exposed to the natural sun. Birds that were allowed the sight of the blue sky through one window only, tended to react as if this window represented the sun and took their directions accordingly. If the incidence of light was changed by the use of mirrors, the direction chosen by the starlings changed correspondingly. When the magnetic field around the experimental chambers was greatly altered by masses of iron, the response of a compass was drastic, but this change in the magnetic field had no effect on the direction of orientation.

In a series of experiments based upon direction learning without the migratory urge, Kramer (1952, 1953) and his co-workers (Kramer & Riese, 1952; Kramer & von St. Paul, 1950) trained starlings to a compass direction in order to be free from the restrictions of the seasons and hours of migration. A circular training cage was used. The birds first learned to go to 1 of 12 containers placed about the circumference for food. The compensation for the sun's motion during the day could be demonstrated by observing that birds which had been trained at a restricted time of day to expect food at a definite angle to the sun chose the same compass direction in critical tests at other times of the day. Similar experiments (Kramer, 1953), but with an artificial sun, have shown that birds alter their absolute direction by changing their angle to the fixed light source. A number of deviations were observed; these errors did not disappear when adjustments of the diameter of the projected sun and the altitude to that of the natural sun were

made. It became evident that the starlings were evaluating only the azimuth, and that it made no difference how high or low the sun was.

In an experiment (Kramer, 1953) with birds raised under sunless conditions, one starling was raised without seeing the sun and trained and tested with a fixed artificial sun. This bird compensated for a clockwise movement it had never experienced. However, when the length of day was adjusted to the fall season and latitude, the bird compensated for a sun-azimuth movement as it occurs in June. All "sunless" starlings raised in 2 subsequent years, however, showed a simple constant angle orientation when tested at times different from training.

The ability of sun-azimuth orientation has been demonstrated experimentally in the starling, red-backed shrike *Lanius collurio,* and warblers (Kramer, 1949, 1950a, 1950b, 1951). In a series of experiments, von St. Paul (1953, 1956) was able to demonstrate orientation in the barred warbler *Sylvia nisoria,* the red-backed shrike, and in the western meadowlark *Sturnella neglecta.* This has also been shown with homing pigeons *Columbia livia* (Kramer, 1957, 1959).

Hamilton (1962b) reared pintails *Dafila acuta,* blue-winged teal *Anas discors,* and green-winged teal *Anas crecca* from eggs laid by wild birds. The eggs were hatched in incubators, and when the ducks were 24 hours old, they were transferred to test cages. To test the ability of ducks to tell direction by the sun, the birds were reared in an environment lacking constant direction save that of the sun and overhead sky. The birds were allowed to obtain water in only one direction, and by the time the birds were 3 weeks old their ability to locate the water source was well developed. In critical tests the experimental birds were removed from the apparatus and deprived of water. In daytime tests with the sun obscured or in twilight periods at dawn and dusk, the birds made random choices, but with the sun directly visible accurate choices were the rule.

The field experiments of Griffin and Goldsmith (1955) showed that common terns *Sterna hirundo* released inland regularly fly southeast for as far as they can be observed. The direction is maintained regardless of the actual home site. When the sky had an overcast, as in previous studies, no orientation could be seen. Similar results were found by Bellrose (1958) with wild mallards *Anas platyrhynchos.* The birds showed a strong tendency to fly north on clear days and nights; the orientation was random with overcast skies. Matthews (1961) observing nonmigratory mallards suggests that the birds upon release flew northwest regardless of the distance and direction of release, season, and length of captivity. Such flights occurred at any time, and with birds of any age, or either sex. These observations suggest sun orientation.

The biological clock which is normally in phase with the astronomical day-night cycle can be controlled and shifted by an artificial light-dark cycle. Starlings studied by Kramer (1957) showed deviation of their clock by producing determinate direction choices. The direction choices were equal to those produced by birds kept in outdoor conditions if the artificial light-dark change was synchronized with the astronomical day-night cycle. They deviated in a predictable manner if the artificial cycle was shifted. A shift of 6 hours resulted in a final directional deviation of roughly 90 degrees. An exposure of 4–12 days to the phase-shifted light-dark cycle was sufficient to attain the full shift of the internal

clock. Other experiments have confirmed these findings (Hoffmann, 1960; Schmidt-Koenig, 1960a).

It is well known that all birds do not migrate during the day. Kramer (1950) suggested the functioning of the sun compass at night. His experiments on directional findings in nocturnal migrants done with Old World shrikes and warblers suggested that orientation was possible only if birds were allowed to see the sun around sunset or earlier. However, Hoffmann's studies (1960) indicated that birds are not able to compensate for the sun's motion during the night and suggested that some other stimulus control was at work.

Studies that do not negate the sun-compass hypothesis but do complicate it have been made by Perdeck (1957), Tinbergen (1956), and Tinbergen and Fijlstra (1954). Perdeck, using Kramer's round cage method, demonstrated that a starling, caught in the autumn, displayed a satisfactory precision in its southwest tendency as long as the sun was visible. However, when the sky became overcast, the bird was at first disoriented but managed to orient again when the sky remained heavily overcast for several days. Observations by Tinbergen of the directional flight of migrating starlings revealed increased scatter on a number of successive cloudy days, but sharp orientation on a cloudy morning following an afternoon when the sun shone for a brief time. This finding would support a celestial clue hypothesis as the oriented flights under overcast can be attributed to a carryover of orientation which was achieved under clearer conditions.

Star-Navigation Hypothesis

It has been noted that warblers along with most small passerines migrate chiefly at night so that the sun is not what guides them. In specially designed cages, Sauer and Sauer (1955) began looking for other factors in orientation. They observed that warblers reared in closed, soundproof chambers under summerlike conditions still got restless at the time when they would have begun their migrations.

Sauer and Sauer (1955) also tested blackcaps *Sylvia atricapilla* and garden warblers *Sylvia borin,* night migrants, for nocturnal migratory directional tendencies. In the autumn migration period, birds were put singly into a circular cage which could be rotated and was screened on all sides so that no surrounding landscape could be seen below an angle of 68 degrees. Definite nocturnal orientation was observed under clear or partly cloudy skies, even if the birds had never seen the sky before the experiment. The fluttering birds kept facing in their natural migration direction, for example, towards the south-southwest and southwest. As long as the brightest stars were still visible through a veil of clouds, the direction of migration orientation was unimpaired. In a diffusely illuminated room, under polarized light, or under a cloudy sky, the birds were completely disoriented. It was noted that the birds showed positive phototropic responses to the moon, bright meteorites, and directed artificial light. In similar experiments during spring migration, blackcaps chose the species-specific direction of migration, north-northeast and northeast.

Further experiments by Sauer and Sauer (1959) with the common whitethroat *Sylvia communis,* garden warbler, wood warbler *Phylloscopus sibilatrix,* and lesser grey shrike *Lanius minor* were made in Southwest Africa, and the birds took up roughly the directions which would be expected. It was suggested that the birds recognized when they had arrived at the end of their

migration by the star pattern. The effect of overcast skies was again stressed. Whitethroats still tended to the south, however, even though they were already 2,000 miles south of their normal winter quarters.

Sauer (1956, 1957a, 1957b) followed his hypothesis of birds' orientation by stars by taking birds into a planetarium. In a Zeiss planetarium, under diffuse illumination, the direction of migration was random. Under an artificial replica of the starlit spring sky a blackcap faced north-northeast and northeast. A lesser whitethroat faced towards the south-southeast and southeast under an artificial autumn sky. An inexperienced lesser whitethroat presented with the starry skies of different latitudes faced between southeast-east in a way which exactly corresponds to the known behavior of experienced birds in nature. The presentation of an autumn sky during spring and vice versa caused the birds to alternate rapidly between the direction of spring and autumn migration. Under a winter or a summer sky during autumn, a lesser whitethroat was completely disoriented. If the constellations were changed to correspond with an eastward or westward displacement of the birds, they were able to compensate partly for this and would have found their way back to their summer or winter quarters. It was suggested that the birds were able to see a part of the star pattern known to them.

A lesser whitethroat (Sauer, 1957b) was presented with an artificial sky as would appear at the latitude of the southern Mediterranean. The bird changed the direction of its movements from southeast to south in the same way as these birds are known to change their course when reaching the eastern end of the Mediterranean, indicating true bicoordinate navigation. It must be noted, however, that such correction was not observed in cases where the westward displacement was simulated.

Orientation of two garden warblers and three blackcaps studied with progressive shifts of the star pattern (forward in time through a total of 24 hours) led to compensation for reorientation, disorientation conflicts between the appropriate migration direction and its reverse, disorientation, and, finally, to normal orientation again. The results were not always consistent and were not duplicated by shifting the pattern backwards in time. Changes in apparent latitude did not seem to affect the conflict type of orientation (Sauer & Sauer, 1960).

In making an evaluation of birds' reactions to simulated changes of position, Kramer (1961) noted that the artificial sky stood still and its time adjustment was consistently retarded in a manner corresponding to a westerly displacement of as much as 15 degrees longitude or more. In addition, the testing apparatus had to be placed excentrically in the planetarium, as the center was occupied by the projector. In general, the cage was placed south of the room center and as the dome was only 6 meters in diameter, parallactic distortions were unavoidable. It must be pointed out that even the displacement of the bird in the cage as it moved must have caused gross visual changes. The net results of these technical problems may have caused marked changes in the behavior of the bird.

Sauer (1961) concluded from his research that birds not only use the stars to determine a fixed compass direction, but also to derive further information, particularly with respect to the seasons, from the temporal arrangements of the constellations. He suggested that various species possess a mechanism of migration orientation which enables them, independent of local topography and

their individual experience, to determine, with the help of their ability to assess time, their species' course of migration while steering by starlit sky. This implies that birds may use a bicoordinate system of star navigation when displaced spatially. But, as yet, it has not been possible to ascertain what celestial features are the principal cues for the bird. In all of Sauer's work the sky zone lower than approximately 35 degrees above the horizon was not shown. If the sky was overcast so that the Milky Way could not be seen but the major stars were yet visible, the bird's orientation was not disrupted. The cues that were used were in all probability of some configurational nature.

The conclusions of Sauer, however, have been complicated by a recent observation of Agron (1962), who has pointed out that Sauer has neglected the evolutionary implications of the relatively rapid changes in the star positions. The stars and seasons reverse their relationships completely every 13,000 years because of the earth's axial precession, and also there are slower changes in star pattern due to "proper motion" of stars. Agron suggested that for birds to navigate by the stars, they must be continuously evolving at a rate which is not incompatible with the ratio of celestial changes.

Using the white-crowned sparrow *Zonotrichia coronata*, Mewald and Gose (1960) observed in an experimental setting that these birds showed strong orientation of nocturnal but not diurnal activity toward the north in the spring and the south in the fall. A more diffuse orientation in the fall was ascribed to the birds being close to their normal wintering area.

Hamilton (1962a) using North Dakota bobolinks *Dolichonyx oryzivorus* under a clear, moonless fall sky in San Francisco reported a remarkable uniformity in preferred direction. The most frequent response under overcast was waning, nondirected response intensity, and subsequent termination of activity; direction was lacking. Response amplitude faded when bright stars were no longer visible. The moon disrupted the directional tendency. When the migratory call notes were recorded and played back, inactive birds could be stimulated to activity and the birds already active to longer than normal activity.

Pintails, blue-winged teal, and green-winged teal under a clear moonless sky consistently made accurate direction choices although some birds could not be induced to respond at all. With the moon in the sky some birds diverted direction of choice toward the moon, others were not affected (Hamilton, 1962b).

Fromme (1961) and Merkel and Fromme (1958), using European robins *Enithacus rubecula*, redbacked shrikes, and whitethroats, observed that the birds preferred a direction in nocturnal migration which corresponded to the natural migratory direction appropriate to the time of year. These observers reported that orientation was just as good without optical points of reference and proposed that these birds must have a means of orientation other than the star pattern shown by Sauer (1956) to be important to other species. They suggested that electromagnetic radiations from celestial bodies could be excluded, however, for investigations (Merkel & Fromme, 1958) with an artificial magnetic field were ambiguous. Migratory direction was completely lost though, when the birds were put into steel chambers. Fromme and Merkel's observations, that birds remained oriented without optical reference points, can not be explained by any of the present the-

ories and are contrary to the findings of others.

Two papers by Wallraff (1960a, 1960c) include detailed criticisms of Sauer's (1956) claims that warblers detect changes in latitude and longitude from corresponding shifts in star patterns. Sauer's data have been recalculated and regrouped with the result that changes in orientation with different latitudes now do appear somewhat less convincing. Changes with differences in longitude are shown to have inconsistencies and are explained as due to the azimuthal compensation of a star compass imposed by a change in time. It must be noted, however, that there is a strong tendency for the grouping of data until very small differences achieve a given level of significance. It is highly conceivable that much of the controversy has its roots well buried in the ground of statistical method.

Sun-Navigation Hypothesis

About the time of Kramer's original work on the sun orientation hypothesis, Matthews in England was studying the critical problem of stimulus control of navigation. This area has been one of the most difficult problems for the student of animal behavior. True navigation involves the ability in animals of "knowing" where they are in unfamiliar surroundings.

Three studies may exemplify this point. Kenyon and Rice (1958) removed 18 adult Layson albatrosses *Diomedea exulans* from nests at Midway Atoll. Fourteen of these returned from widely spaced points of release. The greatest distance of release was 4,120 miles from the nest, and the bird returned in 32 days. Mazzeo (1953) provided an account of a Manx shearwater *Procellaria puffinus* that homed from Boston to its colony off Wales in less than 13 days. This shearwater returned from outside the range of the species along an east-west route at right angles to its normal migration. Recently, Hamilton (1962c) described a report in which a captive bobolink escaped during the fall at Berkeley, California. This bird had been taken from its breeding grounds in North Dakota during the previous summer. The following summer this same bird was recaptured at the location where it had originally been trapped in North Dakota. Since California is not on the migratory pathway of bobolinks, the eventual return of the bird implies some capacity of navigation.

From Matthews' early work (1951, 1953c) the suggestion has come that homing pigeons get their orientation for navigation from the sun. This must be the case, it is argued, for the performance of the birds was not as good under overcast skies. Under this condition the birds scattered randomly when released, and homing was very much better when the sky was clear. The same general findings hold for the Manx shearwater (Matthews, 1953a) and also for gulls (Matthews, 1952). Matthews (1951) has shown that pigeons are able to fly in a given direction when released at unknown points. When released in unknown areas at right angles to, or in the opposite direction from, the original line of training, a good proportion of the birds returned to the loft. It has been emphasized by Matthews (1953b) that homing and initial orientation are two distinct events. Using young and untrained birds, he reported orientation toward the home loft. However, such birds homed poorly. It was thought that practice affected the return rather than the determination of direction bearing. From his studies Matthews hypothesized that homing birds determine latitude on the basis of the sun's altitude and longitude on the basis of time.

Matthews (1955a) stated:

> The essential feature is the sun arc. This is inclined at an angle from the horizontal which is constant for a given place and is a measure of the latitude of that place. . . . The speed at which the sun moves round its arc is, for practical purposes, constant at 15° an hour. When it has reached a particular point on this arc at home, it will have advanced further to an observer in the east, and less far to one in the west, the difference in arc angle (the angle round the arc from the noon position) being directly proportional to the change in longitude. At home the bird will become familiar with the features of the sun arc and the sun's position on it at different (local) times. These will be related to the internal "chronometer" which is also an essential part of the hypothesis. In unfamiliar surroundings the bird will have to construct the sun-arc from observation. The suggestion is that it observes the sun's movement over a small part of its arc and extrapolates to obtain the highest point. Measurement of the altitude of this point, the angle from the horizontal and comparison with the remembered value for home, say, the previous day, will give the latitude change. The arc angle from the observed sun to this highest point when compared with that obtained at home for the same chronometer time will give the longitude change [pp. 92–94].

In support of his assumption of the chronometer, Matthews (1955b) conducted an experiment on that part of the sun-navigation hypothesis which suggests that birds can detect longitude displacement by comparison of home time with local time estimated from the highest point of the sun's arc. Pigeons were exposed for 10 days to an artificial day which was 3 hours in advance of the normal day. When tested, the birds showed no loss in time orientation. Continuing the research, the birds were then exposed to 4 or 5 days of irregular light-dark sequences followed by 11 days of regular sequences, advanced or retarded with respect to normal days. The tests from west, east, and north indicated that the chronometers of the birds had been affected. The pigeons tended to fly in the predicted false direction, east, after an advanced day and west after a retarded one. These results support the general theory. Much of the supporting evidence for sun navigation suggested by Matthews has been criticized, although his studies have not been reproduced by other investigators.

It has been observed that birds show a "position effect," in that they do not find their way equally well from all directions. Kramer (1957); Pratt and Wallraff (1958); Kramer, Pratt, and von St. Paul (1958); and Schmidt-Koenig (1960) have found that birds homed and showed more accurate orientation by flying toward the north than the birds flying toward the south. This northerly orientation may also be related to the postfledgling movements of many birds. This "nonsense orientation" in pigeons, and birds in general, Matthews (1961) argues, has caused confusion in research studies on navigation. The fact that there are directional effects in different geographical localities was thought by Kramer as possibly due to selective differences in the pigeon stock that was used. Hoffmann's (1958) research seems to have eliminated this possibility. Using both English and German stock in his experiments, Hoffmann found no differences between the two. The differences from the original studies, it was suggested, seemed to be due to some unknown aspect of the locality. Wallraff (1960c), in studying the relationship between homing ability of pigeons and meteorological and geophysical factors, found that low wind velocities and clouding had a negligible effect on homing ability, though air temperatures and barometric pressures had a great influence. In time, it may be discovered that the position effect can be explained by a critical analysis of the homing locale. However, Kramer (1961) noted day-to-day variations in

homing success which could not be explained in terms of traditional meteorology.

Homing ability also shows marked seasonal fluctuations. Kramer and von St. Paul (1956) and Kramer (1957, 1959) observed that homing was much poorer in winter than in summer at least in Northern Germany. The very poor results obtained in winter over very short distances showed that the birds did not use visual information from the surroundings.

Kramer (1959) and Wallraff (1959) have studied orientation independent of landmarks. Pigeons which were raised in open aviaries exhibited homeward orientation when released. No such orientation was displayed by birds which were raised in walled aviaries allowing vision only of the sky down to 30 degrees above the horizon, in aviaries allowing a view of a sector of the northern sky and horizon, or in aviaries surrounded by palisades in such a manner that free sight of the sky was not allowed except for a zone of elevation 3 degrees above the horizon. Using the bank swallow *Chaetura pelogica* Sargent (1962) tested birds in a portable orientation cage. The birds showed preferences of direction at distances up to 25 miles provided a view of surroundings was available to them during transit, but showed no such tendencies at greater distances or if not permitted to observe surroundings during transit.

The sun-arc navigation theory requires observation of the sun's movement for extrapolation. Kramer (1953) denied pigeons the observation of the sun's path. He demonstrated that the birds quickly oriented themselves even if they were released at the time of day when the elevation of the sun, at the release site, was the same as that of the home loft.

In testing aspects of Matthews' theory, Pratt and Thouless (1955) allowed their pigeons to see the sun for varying periods of time before release, while other birds were screened until release. They observed that both groups oriented towards home with equal accuracy and with such prompt headings after release that there was some question that the birds were judging the trend of the sun's arc. Pratt (1955) and Pratt and Thouless (1955), using both trained and untrained birds, found that both groups showed a significant tendency to fly in the homeward section of the release circle. Likewise, both groups homed with about the same speed and success. This similarity was contrary to the results obtained by Matthews. Like Kramer, Pratt (1955) attributed the differences in results to stock differences. It has also been pointed out that another possible interpretation lies in the fact that Matthews did not give free range to his birds in the loft, whereas Pratt allowed such activity. The restricted birds, upon arrival in the vicinity of the loft, may not have been able to recognize their home territory.

Hoffmann (1954, 1958, 1960) and Schmidt-Koenig (1960a, 1960b) reported a number of findings with pigeons whose biological clocks were shifted. These data showed that the sun was definitely involved in orientation, but only to the extent that it served as a compass. In the sun-arc theory, following a shift of the bird's chronometers 6 hours clockwise, only at one particular time of day, at 9 A.M., do the altitudes of the anticipated and realized sun agree. All birds that were released before that time should have taken their heading toward the east to southeast, and all birds released after that time toward the north to northwest. This was clearly not the case for there was a deviation to the left with reference to the headings of the control birds, no

matter what time and from which direction or distance the birds were released. This evidence was more in keeping with an orientation reaction rather than true navigation. Wallraff (1960a), after summarizing the literature on navigation, concluded by stating:

> It has not yet been possible to demonstrate navigation on a purely astronomical basis in any animal. At the present the hypothesis of sun navigation is to be regarded as disproved, at least in pigeon homing, that of star navigation as not proven [p. 459].

Whether migratory flight is governed by goal or directional orientation has been the subject of a whole series of field displacement experiments with the European starling (Kratzig & Schuz, 1936; Perdeck, 1954, 1957), hooded crow *Corvus cornix* (Ruppell, 1944), American crow *Corvus brachyrhynchos* (Rowan, 1946), European sparrow hawk *Accipter nisus* (Dorst, 1939), and European white stork *Ciconia ciconia* (Schuz, 1949, 1950, 1951; Thienemann, 1931). In all of the tests, birds of the same year flew parallel to the normal migratory route; whereas older birds displayed a stronger tendency to home towards the normal distribution area of their population.

Perdeck (1958) suggested that the birds possess at least two different mechanisms for orientation. He observed on the basis of recovery of banded birds that juvenile starlings displaced transversely to their normal route before or during their first migration, flew in the original direction after displacement. He concluded that what they used for orientation must have been a sort of compass to find their direction. Adult birds headed straight to their wintering areas if displaced, though they flew in directions that differed from their normal migratory course over long areas unknown to them.

There seems to be adequate support for the conclusion that there are great differences among species in the capacity to navigate. In surveying many tests of homing ability in the nonmigratory titmice *Paridae,* Hinde (1952) concluded that these birds can find their way home up to a maximum of 7 miles. The coming into home territory seemingly determines the rate of return. In a similar study, Wojtusiak (1949), experimenting with the European tree sparrow *Passer montanus,* also nonmigratory, found no birds returned home from distances greater than 7 miles. The extraordinary abilities in certain migratory sea birds, have, on the other hand, been demonstrated. Lack and Lockley (1938) have reported Manx shearwaters removed from their nesting borrows returned successfully to the colony from points as remote as 930 miles and from areas outside their normal range.

Pennycuick (1960a) has criticized the sun-arc hypothesis from a theoretical point of view. He suggested that a change of azimuth can only be measured by reference to some object which does not move sideways. A bird, then, could only make the measurement of a change in azimuth by reference to a fixed object when the bird is stationary on the ground. Over the sea it is not physically possible to measure it at all without some kind of compass. However, he pointed out that birds do navigate over water. Therefore, Pennycuick saw Matthews' theory as untenable on the grounds of the physical impossibility of making one of the measurements. When Matthews considered errors of position, he assumed that they would appear as errors in latitude. Pennycuick questioned how meridian altitude data could be calculated by Matthews' theory.

Pennycuick's arguments were shown

by Pumphrey (1960) to be fallacious. What the bird needs, according to Pumphrey, is an indication of the direction on the horizontal plane long enough to allow the change of azimuth. It may be assumed that a flying bird could obtain the necessary information in flights over the sea, and does not need a fixed object. It was conceded by Pennycuick (1960b) that the rate of change of the sun's azimuth could, at least in principle, be measured by possible reference to, and from analysis of, the waves on the sea. Wave patterns are, under some conditions, more or less regular, but in general seem to be too irregular to be used for measurement of sun angles.

It can be held that Matthews' theoretical position has not been proven in that there is currently little evidence in its support. Nonetheless, there is no doubt that given species do navigate over great distances, and some form of bicoordinate sun navigation cannot be ruled out as the method.

A simpler sun-navigation hypothesis has in turn been suggested by Pennycuick (1960a). It was supposed that upon release the bird measures the sun's elevation and its rate of change and compares these with the conditions at home. Schmidt-Koenig (1961) commented on Pennycuick's theory and argued that the experimental evidence from previous research does not agree with the theory. Under given aspects of a phase shift of the chronometer, birds came to expect a decrease in altitude but did in fact observe an increase in altitude. Instead of an eastern heading, as would be predicted by the theory, there was a left deviation.

From Pennycuick's theory of sun navigation follows the possibility of the use of stars for orientation. Theoretically the use of stars for navigation should be easier than use of the sun. He has pointed out that the altitudes of two different stars can be used simultaneously as two referent points and that there is no need to measure the rates of change, as viewed from Sauer's planetarium experiments in which the star pattern did not move.

Furthermore, it is a well-known phenomenon that under given situations the human observer can detect some of the brighter stars during the day, and if, by chance, the bird is capable of such observations under "normal" circumstances, a new much more simple navigation theory would be possible. At the present time it is not possible to make any reliable statement about the discriminative basis for navigation. It has been conjectured that vision is responsible, but this can scarcely be more than a generalization as no data are available for the quantities the bird is supposed to be measuring.

Other Attempts to Explain True Navigation

Wojtusiak (1946, 1949), reporting on studies made in Poland, brought forth the theory that birds may be sensitive to the infrared visual spectrum. If the theory holds, birds could see through fog and also at night. There is no satisfactory experimental support that birds use the infrared frequencies (Matthews & Matthews, 1939). Watson and Lashley (1915), Lashley (1916), Hecht and Pirenne (1940), and recently Blough (1957) have evidence that birds' visual spectrum is approximately that of man, and therefore, there appears to be no evidence for the theory.

Ising (1946) proposed that the navigation of birds may depend upon the Coriolis force. The Coriolis force is a mechanical effect which results from the rotation of the earth; the strongest force is at the poles and is zero at the equator. As a result of the rotation of

the earth, a flying bird's weight changes depending on the direction of flight and its latitude. Theoretically a bird could judge latitude. Wynne-Edwards (1948) has shown, however, that a change of speed from 40 to 39 miles per hour in the bird's flight could alter the Coriolis effect by 2.5%, and this same change would correspond to a displacement of about 150 miles. The Coriolis force is very small and could be offset by many confusing circumstances such as the effect of variable winds and air turbulences. The measurement of the changes that are required by the theory seems outside the organ sensitivity of the bird (de Vries, 1948; Griffin, 1955; Thorpe & Wilkinson, 1946).

Since the nineteenth century, various theories of navigation based on the geomagnetic field have been in circulation, but as yet no one has produced experimental evidence that any animal is sensitive to the magnetic field. Nonetheless, Yeagley (1947, 1951) reported accounts of his experiments with homing pigeons designed to show use of both magnetism and the Coriolis force. He suggested that birds determine the latitude by the effects of the Coriolis force and the longitude from the variations in strength of the earth's magnetic field. The variation in these two forces may be thought of as constituting a grid in North America; one set of lines, equaling the Coriolis force, being the parallels of latitude, and the other being arcs drawn about the North Magnetic Pole as a center. He reported an experiment with pigeons that had been trained to return to a loft in Pennsylvania. These birds were then released in Nebraska where the lines of the two forces intersect as they do at the loft in Pennsylvania. In actuality only a very few birds even reached the general area of the Nebraska loft. However, Yeagley considers that the results showed a significant tendency to fly in that direction. Yeagley also reported positive evidence for his position when comparing homing abilities of two groups, one with magnets attached to the wings, which was to prevent any measurement of magnetism, and a second group with unmagnetized copper bars attached in similar fashion. It was stated that the magnets produced a magnetic field of about the same intensity as the vertical component of the earth's field. Birds without magnets returned more rapidly and in greater numbers. On subsequent tests Yeagley failed to find any differences between the groups. This was explained on the basis of magnetic disturbances occurring on the testing day.

Several other investigators have attempted to replicate the experiment but have found no significant differences between birds with and without magnets (Matthews, 1951; Van Riper & Kalmbach, 1952). Griffin (1955), examining the Yeagley study, has pointed out differences in method of attachment of the magnets and copper and has shown that several magnets dropped off the birds in one group while all the metal remained attached in the control group. Differences in attachment may have resulted in differences in irritation and injury. Griffin (1955) and Orgel and Smith (1954) reported negative outcomes of attempts to elicit given responses to a magnetic field in the laboratory. It has also been shown that responses of orientation were not affected (Fromme, 1961; Kramer, 1950a, 1950b). Wynne-Edwards (1948), Thorpe (1949), Griffin (1952, 1955), and Matthews (1955) have presented fundamental criticisms of Yeagley's work. Yeagley's theory appears disproved because it has not withstood critical replication.

However, the theory may not be completely dismissed. It has been reported

by Stewart (1957) that electric conditions are created by air friction on the moving bird's feathers. The magnetic field surrounding it cannot help but react to the earth's magnetic fields. This in turn could cause the tips of the wings to move, and thus, the bird may detect this movement. It is thought by Stewart that a bird in flight could correct its direction from these magnetic cues. Nonetheless, evidence for orientation and navigation without the use of celestial clues is not very convincing.

REFERENCES

Agron, S. L. Evolution of bird navigation and the earth's axial precession. *Evolution,* 1962, **16**, 524–527.

Bellrose, F. C. Celestial orientation by wild mallards. *Bird Band.,* 1958, **29**, 75–90.

Blough, D. S. Spectral sensitivity in the pigeon. *J. Opt. Soc. Amer.,* 1957, **47**, 827–833.

de Vries, H. L. Die Reizschwelle der Sinnesorgane als physikalisches Problem. *Experientia,* 1948, **4**, 205–213.

Dorst, R. Über den Einfluss von Verfrachtungen zur Herbstzugzeit auf den Sperber, *Accipiter nisus* (L.). *Proc. 9th Int. Ornithol. Congr. Rouen,* 1939, **100**, 503–521.

Fromme, H. G. Untersuchungen über das Orientierungsvermögen nächtlich ziehender Kleinvögel. *Z. Tierpsychol.,* 1961, **18**, 205–220.

Griffin, D. R. Bird navigation. *Biol. Rev.,* 1952, **27**, 359–393.

Griffin, D. R. Bird navigation. In A. Wolfson (Ed.), *Recent studies in avian biology.* Urbana: Univer. Illinois Press, 1955. Pp. 154–197.

Griffin, D. R., Goldsmith, T. H. Initial flight directions of homing birds. *Biol. Bull.,* 1955, **108**, 264–274.

Hamilton, W. J., III. Bobolink migratory pathways and their experimental analysis under night skies. *Auk,* 1962, **79**, 208–233. (a)

Hamilton, W. J., III. Celestial orientation in juvenal waterfowl. *Condor,* 1962, **64**, 19–33. (b)

Hamilton, W. J., III. Does the bobolink navigate? *Wilson Bull.,* 1962, **74**, 357–366. (c)

Hecht, S., & Pirenn, M. H. The sensibility of the nocturnal long-eared owl in the spectrum. *J. gen. Physiol.,* 1940, **23**, 709–717.

Hinde, R. A. The behavior of the great tit (*Parus major*) and some other related species. *Behaviour,* 1952, Suppl. No. 2.

Hoffmann, K. Versuche zu der im Richtungsfinden der Vögel enthaltenen Zeitschätzung. *Z. Tierpsychol.,* 1954, **11**, 453–475.

Hoffmann, K. Repetition of an experiment on bird orientation. *Nature, London,* 1958, **181**, 1435–1437.

Hoffmann, K. Experimental manipulation of the orientational clock in birds. *Cold Spr. Harb. Symp. quant. Biol.,* 1960, **25**, 379–387.

Ising, G. Die physikalisch Möglichkeit eines tierischen Orientierungssinnes auf Basis der Erdrotation. *Ark. Mat. Astron. Fys.,* 1946, **32A**(No. 18).

Kenyon, K. W., & Rice, T. W. Homing of the Layson albatross. *Condor,* 1958, **60**, 3–6.

Kramer, G. Über Richtungstendenzen bei der nächtlichen Zugunruhe gekäfigter Vögel. In E. Mayr & E. Schüz (Ed.), *Ornithologie als Biologische Wissenschaft.* Heidelberg: Winter, 1949. Pp. 269–283.

Kramer, G. Orientierte Zugaktivität gekäfigter Singvögel. *Naturwissenschaften,* 1950, **37**, 188. (a)

Kramer, G. Weitere Analyse der Faktoren, welche die Zugaktivität des gekäfigten Vögels orientieren. *Naturwissenschaften,* 1950, **37**, 377–378. (b)

Kramer, G. Eine neue Methode zu Erforschung der Zugorientierung und die bisher damit erzielten Ergebnisse. *Proc. 10th Int. Ornithol. Congr. Uppsala,* 1951, 269–280.

Kramer, G. Experiments on bird navigation. *Ibis,* 1952, **94**, 265–285.

Kramer, G. Die Sonnenorientierung der Vögel. *Verh. zool. Gesellsch. Frieburg,* 1953, 72–84.

Kramer, G. Experiments on bird orientation and their interpretation. *Ibis,* 1957, **99**, 196–227.

Kramer, G. Recent experiments on bird orientation. *Ibis,* 1959, **100**, 399–416.

Kramer, G. Long-distance orientation. In A. J. Marshall (Ed.), *Biology and comparative physiology of birds.* Vol. 2. New York: Academic Press, 1961. Pp. 341–371.

Kramer, G., Pratt, J. G., & von St. Paul, U. Neue Untersuchungen über den "Richtungseffekt." *J. Ornithol.,* 1958, **99**, 178–191.

Kramer, G., & Riese, E. Die Dressur von Brieftauben auf Kompassrichtung in Wahlkäfig. *Z. Tierpsychol.,* 1952, **9**, 245–251.

Kramer, G., & von St. Paul, U. Stare (*Sturhus vulgaris*) lassen sich auf Himmelsrichtungen dressieren. *Naturwissenschaften,* 1950, **37**, 526–527.

Kramer, G., & von St. Paul, U. Weitere Erfahrungen über den "Wintereffekt" beim Heimfindvermögen von Brieftauben. *J. Ornithol.,* 1956, **97**, 353–370.

Krätzig, H., & Schüz, E. Ergebnis der Versetzung Ostbaltischer Stare ins Binnenland. *Vogelzug,* 1936, **7**, 163–175.

Lack, D., & Lockley, R. M. Stokholm Bird Observatory homing experiments: I. 1963–37. Puffins, storm-petrels and Manx shearwaters. *Brit. Birds,* 1938, **31**, 242–248.

Lashley, K. S. The color vision of birds: I. The spectrum of the domestic fowl. *J. anim. Behav.,* 1916, **6**, 1–26.

Matthews, G. V. T. The experimental investigation of navigation in homing pigeons. *J. exp. Biol.,* 1951, **28**, 508–535.

Matthews, G. V. T. An investigation of homing ability in two species of gulls. *Ibis,* 1952, **94**, 243–264.

Matthews, G. V. T. Navigation in the Manx shearwater. *J. exp. Biol.,* 1953, **30**, 370–396. (a)

Matthews, G. V. T. The orientation of untrained pigeons: A dichotomy in the homing process. *J. exp. Biol.,* 1953, **30**, 268–276. (b)

Matthews, G. V. T. Sun navigation in homing pigeons. *J. exp. Biol.,* 1953, **30**, 243–267. (c)

Matthews, G. V. T. *Bird navigation.* Cambridge: University Press, 1955. (a)

Matthews, G. V. T. An investigation of the "chronometer" factor in bird navigation. *J. exp. Biol.,* 1955, **32**, 39–58. (b)

Matthews, G. V. T. "Nonsense" orientation in mallard *Anas Phatyhnchas* and its relation to experiments on bird navigation. *Ibis,* 1961, **103**, 211–230.

Matthews, L. H., & Matthews, B. H. C. Owls and infra-red radiation. *Nature, London,* 1939, **143**, 983.

Mazzeo, R. Homing of the Manx shearwater. *Auk,* 1953, **70**, 200–201.

Merkel, R. W., & Fromme, H. G. Untersuchungen über das Orientierungsvermögen nächtich ziehender Rotkelchen. *Naturwissenschaften,* 1958, **45**, 499–500

Mewald, L. D., & Gose, R. G. Orientation of migratory restlessness in the white-crowned sparrow. *Science,* 1960, **131**, 105–106.

Orgel, A. R., & Smith, J. C. Test of the magnetic theory of homing. *Science,* 1954, **120**, 891–892.

Pennycuick, C. J. The physical basis of astro-navigation in birds: Theoretical consideration. *J. exp. Biol.,* 1960, **37**, 573–593. (a)

Pennycuick, C. J. Sun navigation by birds. *Nature, London,* 1960, **188**, 1127–1128. (b)

Perdeck, A. C. Jaarverslag 1953. *Jversl. Vogeltrekst. Texel,* 1954, 3–13.

Perdeck, A. C. Stichting Vogeltrekstation Texel Jaarverslag over 1956. *Limosa,* 1957, **30**, 62–75.

Perdeck, A. C. Two types of orientation in migrating starlings, *Sturnus Vulgaris* L., and chaffinches, *Fringilla coolebs* L., as revealed by displacement experiments. *Ardea,* 1958, **46**, 1–37.

Pumphrey, R. J. Sun navigation by birds. *Nature, London,* 1960, **188**, 1127.

Pratt, J. G. An investigation of homing ability in pigeons without previous homing experience. *J. exp. Biol.,* 1955, **32**, 70–73.

Pratt, J. G., & Thouless, R. H. Homing orientation in pigeons in relation to opportunity to observe the sun before release. *J. exp. Biol.,* 1955, **32**, 140–157.

Pratt, J. G., & Wallraff, H. G. Zweirichtungs-Versuche mit Brieftauben: Langstreckenflüge auf der Nord-Süd Achse in Westdeutschland. *Z. Tierpsychol.,* 1958, **15**, 332–337.

Rowan, W. Experiments in bird navigation. *Trans. Roy. Soc. Canada,* 1946, **40**, 123–235.

Ruppell, W. Versuche über das Heimfinden ziehender Nebelkrähen nach Verfrachtung. *J. Ornithol.,* 1944, **92**, 106–132.

Sargent, T. A study of homing in the bank swallow. *Auk,* 1962, **79**, 234–246.

Sauer, E. G. F. Further studies on stellar orientation of nocturnally migrating birds. *Psychol. Forsch.,* 1961, **24**, 224–244.

Sauer, E. G. F., & Sauer, E. M. Star navigation of nocturnal migrating birds. *Cold Spr. Harb. Symp. quant. Biol.,* 1960, **25**, 463–473.

Sauer, F. Zugorientierung einer Mönchsgrasmücke (*Sylvia a. atricapilla,* L.) unter künstlichem Sternenhimmel. *Naturwissenschaften,* 1956, **53**, 231–232.

Sauer, F. Astronavigatorische Orientierung einer unter künstlichem Sternenhimmel verfrachteten Klappergrasmücke, *Sylvia C. currca* (L.). *Naturwissenschaften,* 1957, **44**, 71. (a)

Sauer, F. Die Sternenorientierung nächtlich ziehender Grasmücken (*Sylvia atricapilla, borin* und *curruca*). *Z. Tierpsychol.,* 1957, **14**, 29–70. (b)

SAUER, F., & SAUER, E. Zur Frage der nächtlichen Zugorientierung von Grasmücken. *Rev. Suisse Zool.*, 1955, **62**, 250–259.

SAUER, F., & SAUER, E. Nächtliche Zugorientierung europäischer Zugvögel in Südwestafrika. *Vogelwarte*, 1959, **20**, 4–31.

SCHMIDT-KOENIG, K. Internal clocks and homing. *Cold Spr. Harb. Symp. quant. Biol.*, 1960, **25**, 389–393. (a)

SCHMIDT-KOENIG, K. The sun-azimuth compass: One factor in the orientation of homing pigeons. *Science*, 1960, **131**, 826–828. (b)

SCHMIDT-KOENIG, K. Sun navigation in birds. *Nature, London*, 1961, **190**, 1025–1026.

SCHÜZ, E. Die Spätauflassung ostpreussischer Jungstörche in Westdeutschland 1933. *Vogelwarte*, 1949, **15**, 63–78.

SCHÜZ, E. Früh Auflassung ostpreussischer Jungstörche in Westdeutschland durch die Vogelwarte Rossitten 1933-1936. *Bonn. zool. Beitr.*, 1950, **1**, 239–253.

SCHÜZ, E. Überblick über die Orientierungsversuche der Vogelwarte Rossitten (jetz: Vogelwarte Radolfzug). *Proc. 10th Int. Ornithol. Congr. Upsala*, 1951, 249–268.

STEWART, O. J. A birds inborn navigational device. *Trans. Ky. Acad. Sci.*, 1957, **18**, 78–84.

THIENEMANN, J. Von Vogelzug in Rossitten. *J. Neumanns, Neudamm*, 1931.

THORPE, W. H. A discussion on the orientation of birds on migratory and homing flights: Recent biological evidence for the methods of bird orientation. *Proc. Linn. Soc. Lond.*, 1949, **160**, 85–94.

THORPE, W. H., & WILKINSON, D. H. Ising's theory of bird orientation. *Nature, London*, 1946, **158**, 903–904.

TINBERGEN, N. Field observations of migration and their significance for the problems of migration. *Ardea*, 1956, **44**, 231.

TINBERGEN, N., & FIJLSTRA, J. J. De veldwaarnemingen: Een schakel die ontrak. *Jversl. Vogeltrekst. Texel*, 1954, 14–22.

VAN RIPER, W., & KALMBACH, E. R. Homing not hindered by wing magnets. *Science*, 1952, **115**, 577–578.

VON ST. PAUL, U. Nachweis der Sonnenorientierunger bei nächtlich ziehenden Vögeln. *Behaviour*, 1953, **6**, 1–7.

VON ST. PAUL, U. Compass directional training of western meadowlarks (*Sturnella neglecta*). *Auk*, 1956, **73**, 203–209.

WALLRAFF, H. G. Örtlich und zeitlich bedingte Variabilität Heimkehr verhaltens von Brieftauben. *Z. Tierpsychol.*, 1959, **16**, 513–544.

WALLRAFF, H. G. Does celestial navigation exist in animals? *Cold Spr. Harb. Symp. quant. Biol.*, 1960, **25**, 451–461. (a)

WALLRAFF, H. G. Können Grasmücken mit Hilfe des Sternenhimmels navigeren? (Kritische Bearbeitung einiger Planetariumsversuche von F. Sauer.) *Z. Tierpsychol.*, 1960, **17**, 165–177. (b)

WALLRAFF, H. G. Über Zusammenhänge des Heimkehrverhaltens von Brieftauben mit meteorologischen und geophysikalischen Faktoren. *Z. Tierpsychol.*, 1960, **17**, 82–113. (c)

WATSON, J. B., & LASHLEY, K. S. Homing and related activities of birds. *Pap. Tortugas Lab.*, 1915, **7**, 1–104.

WOJTUSIAK, R. J. Hypothesis of sensibility to infra-red rays as an attempt to explain some problems of orientation in animals. *CR Sci. Math. Nat. Acad. Polon.*, 1946, 28–29.

WOJTUSIAK, R. J. Polish investigations on homing in birds and their orientation in space. *Proc. Linn. Soc. Lond.*, 1949, **160**, 99–108.

WYNNE-EDWARDS, V. C. Yeagley's theory of bird navigation. *Ibis*, 1948, **90**, 606–611.

YEAGLEY, H. L. A preliminary study of a physical basis of bird navigation. *J. appl. Physiol.*, 1947, **18**, 1035–1063.

YEAGLEY, H. L. A preliminary study of a physical basis of bird navigation. Part 2. *J. appl. Physiol.*, 1951, **22**, 746–760.

(Received October 21, 1963)

Psychological Bulletin
1964, Vol. 62, No. 3, 180–196

EXAMINATION OF GIBSON'S PSYCHOPHYSICAL HYPOTHESIS [1]

WILLIAM EPSTEIN AND JOHN PARK

University of Kansas

The experimental evidence regarding Gibson's psychophysical hypothesis is examined. The presentation is divided into 2 sections dealing with static and transforming stimulation, respectively. Under the former heading the stimulation for surface, slant, and depth is considered. Under the latter heading the investigations of motion perspective, continuous perspective transformations, and size-transformations are examined. 2 general conclusions emerged from this examination: the psychophysical hypothesis has not been unequivocally confirmed and Gibson's theory requires revision to make explicit several recurrent implicit premises.

One of the major contemporary theoretical accounts of perception is Gibson's (1950b, 1959, 1963) psychophysical theory. Unlike other theories, for example, Gestalt and transactionalism, Gibson has attempted to develop a S-R account of perception; one which does not introduce intervening variables such as sensory organization or neo-Helmholtzian assumptive worlds. The central proposition of Gibson's (1959) theory is the generalized psychophysical hypothesis:

> for every aspect or property of the phenomenal world of an individual in contact with his environment, however subtle, there is a variable of the energy flux at his receptors, however complex, with which the phenomenal property would correspond if a psychophysical experiment could be performed [p. 465].

The objective of this paper is to evaluate the evidence for this hypothesis.

The similarity between Gibson's hypothesis and the older discredited constancy hypothesis is evident. Gibson's hypothesis, however, is less vulnerable. Unlike the earlier hypothesis, Gibson's position does not presume a picture theory of the proximal-perceptual relation. Gibson requires retinal correlates, not retinal replicas. A more important source of strength is Gibson's conceptualization of the stimulus. In the earlier view the stimulus was a static, nonchanging, local correspondent of the distal source. As long as this view prevailed it was a simple matter to contradict the constancy hypothesis by demonstrating the equivocality of the distal-proximal and proximal-perceptual relationships (Koffka, 1935, pp. 75–105). This type of critical analysis is not so readily applicable when Gibson's (1959, 1960) view of the stimulus is taken into account. In Gibson's (1959) view the effective stimulation for perception must be sought "in a textured optical array, supplemented by the transformations relating a simultaneous pair of them, and by the transformations relating a sequence of momentary arrays [p. 474]." The stimulus is ordinal—a pattern or gradient of stimulation. The order may be sequential as well as spatial. Sequential ordinal stimulation refers to the lawful transformations of stimulation which occur when either the subject or the object is in

[1] This article was written with the support of Research Grant MH 04153-04 to the first author by the National Institute of Mental Health of the United States Public Health Service.

motion. It is important to note that in Gibson's view (see especially 1957 and Gibson & Gibson, 1957) the change in stimulation is a stimulus to which the subject can respond. On other occasions, the subject may respond to the aspect of stimulation which remains invariant in the flux of stimulus transformations. The invariants uniquely specify the source of stimulation, and the perceptual apparatus is geared to extract these invariants, thereby achieving veridical perception. Thus, in Gibson's view, stimulation undergoing change is more informative than static stimulation since an instance of static nontransformation can rarely specify its source. Gibson's conceptualization of stimulation provides him with the grounds for rejecting the frequent assertion that the distal-proximal and proximal-perceptual relationships are fundamentally ambiguous. Once this assertion can be denied the way is opened for a retinal account of perception.

In one sense the greater portion of all experimentation on perception could be considered to have bearing on Gibson's hypothesis. However, our treatment will be restricted mainly to those studies having the following two characteristics: (*a*) The main objective of the experiment was to examine psychophysical correspondence. (*b*) The experiment has sought to manipulate the type of visual input, that is, optic array, which Gibson has stressed. For this reason we have excluded discussion of other studies which might be considered relevant, for example, the work reviewed by Epstein, Park, and Casey (1961) and Epstein and Park (1963). The discussion will be divided into two parts, dealing with static and transforming stimulation, respectively.

Psychophysical Correspondence under Conditions of Static, Nontransforming Stimulation

This section will review the experimental studies of the perception of surface, slant, and distance.

Stimulation for the Perception of Surface

"According to what will here be called the *texture-hypothesis,* the stimulus for a visual surface is a fully differentiated, sharp, or textured retinal image [Gibson & Dibble, 1952, p. 414]." Evidence for this hypothesis is adduced by Gibson from a series of experiments with the *Ganzfeld.* In an experiment similar to Metzger's (1930) prototypal research, Gibson and Dibble (1952) seated the subject before a gray wall of coarse-textured plasterboard which filled nearly the entire visual field. With full illumination, the wall appeared surfacelike and visibly textured. As the illumination of the wall was gradually reduced, the texture vanished, and the impression of hardness was replaced by that of a fog. In a similar manner the impression of surface vanished when Gibson and Waddell (1952) produced homogeneous visual stimulation by means of translucent eyecaps. These results are in accord with the findings of other investigators (Cohen, 1957; Hochberg, Triebel, & Seaman, 1951; Metzger, 1930) and appear to support the hypothesis that the perceived hardness of a surface is related to the steepness of the gradients of luminous intensity in the retinal image.

While it appears reasonably certain that texture is sufficient for the perception of a surface, there is evidence that it is not necessary. Gibson and Dibble (1952) found that a homogeneous tex-

tureless surface viewed through a small aperture appeared to be a hard surface in the plane of the reduction screen rather than a film in the aperture. The authors suggest that the effective stimulus may have been the steep intensity gradient at the contour. Cohen (1957) found that a steep gradient of intensity between a homogeneous disc and a homogeneous field produced the impression of a figure having a hard surface and lying on a foggy ground.

Stimulation for the Perception of Slant

Gibson (1950a) proposed that "slant at any point may be given by the rate of increase of density of the texture at that point [p. 371]." This hypothesis has had to be revised in the face of evidence that rate of increase of texture density is not a sufficient stimulus for the veridical perception of slant. Since a gradient of texture density merely specifies a family of tridimensional arrangements, it is understandable that the corresponding percept may be variable. In addition, slant is generally underestimated when texture gradient is the chief or only cue.

Underestimation has been found by Gibson (1950a), Gruber and Clark (1956), and Clark, Smith, and Rabe (1956). The stimuli were textured surfaces presented at several degrees of slant, or projections of lantern slides made by photographing textures at several slants. The textures employed were various: bricklike, mottled, or composed of irregular light dots on dark grounds. Outline perspective, binocular cues, and motion parallax were eliminated. Perceived slant increased as the steepness of texture gradient increased. Hence it was concluded that a retinal gradient of texture density can, in isolation from other factors, determine perceived slant. Nevertheless, slant was underestimated (see Gruber & Clark, 1956). Although the amount of underestimation was considerably less, some underestimation occurred even for regular textures consisting of repeated evenly-distributed elements with linear perspective and familiar shape. These results may have been caused in part by the tendency of perception to conform to the slant of the reduction screen.

These experiments failed to observe the distinction between optical slant and geographical slant. Optical slant is dependent only on the geometrical relation of a surface to the eye, whereas geographical slant is dependent on the relation of the surface to other parts of the world or to gravity. By creating an incongruency between the reference-axes of the eye itself and the reference-axes of their experimental room, Gibson and Cornsweet (1952) produced a situation in which the two kinds of slant could be perceived separately. The subject's head was rotated 45 degrees to the left around its vertical axis and set at a 45-degree angle to the walls of the room. The subjects were able to determine accurately the point at which a rotating textured surface reached each of two normal positions: perpendicular to the line of sight (a judgment of optical slant) or parallel to the walls of the room (a judgment of geographical slant). The results indicate that the two kinds of slant can be perceived independently, and that optical slant corresponds to the gradient of density of texture at the fovea, while geographical slant does not.

Gibson, Purdy, and Lawrence (1955) investigated the role of texture density by means of an optical tunnel, a device composed of a set of plastic sheets, alternately black and white, with circular apertures, set one behind the other so as to project onto the retina a set of concentric rings of alternating high and low intensity. When subjects with motionless heads and monocular vision viewed a tunnel having a zero gradient of density, they reported seeing something ambiguous, something which fluc-

tuated between flat and deep. When subjects were given another presentation of this same pseudotunnel using both eyes and a third presentation using both eyes without a biting board, they reported a surface with extended depth which approached the actual depth of the pseudotunnel. When the peripheral-to-central gradient of disparity was reversed by a pseudoscope, the subjects saw a striped, convex, truncated cone protruding from a background. It appears that an unambiguous appearance of slant results from a zero gradient of density only when it is supplemented by some other stimulus such as a gradient of disparity or deformation. Gibson, Purdy, and Lawrence (1955) concluded that a gradient of texture density cannot by itself compel a perception of corresponding slant because the gradient merely specifies a family of tridimensional arrangements rather than a particular distal stimulus situation. For example, a gradient of increasing texture density may be produced either by a slanted surface having equal-sized elements at constant distances from one another or by a frontal-parallel surface having elements with progressively diminishing sizes and separations. What a given gradient of density determines, according to Gibson, is "a relationship between apparent slant and apparent spacing of texture elements [Gibson et al., 1955, p. 12]."

What little evidence there is regarding this hypothesis comes from an experiment by Beck (1960). He presented a set of optical tunnels which had constant-sized apertures and elements unevenly spaced so as to duplicate the increasing (or decreasing) density gradients which would be present in the cross sections of rays projected by equally-striped, tapered tunnels (or cones) converging to various angles. With fixed monocular vision, these texture gradients produced impressions of noncylindrical tapered objects. However, these impressions were quite variable. Hence it appears that a texture gradient in isolation is able to produce a complex impression of slant and recession, but it is not able to uniquely determine slant and recession. A particular impression of slant and recession presupposed "a presumption, attitude, or expectation," for example, that the stripes along the surface of the pseudotunnel were spaced equally. A linkage between apparent slant and apparent texture-spacing is indicated by the fact that adding binocular parallax resulted in a change of subjects' judgments from an equally striped, tapered tunnel to an unequally striped, parallel tunnel. However, some exceptions were found to the proposed linkage between apparent slant and apparent texture-spacing. Beck also found that the introduction of binocular disparity did not always eliminate the ambiguity of fixed monocular vision.

That texture gradient is not adequate as a cue to slant is also confirmed by Smith and Smith (1957), who required subjects to make judgments of the degree of curvature of a perfect semicylinder with a textured surface and hidden ends. Under monocular conditions the mean judgments were nearly flat. This result was explained by the authors as due to the tendency to see the display in the plane of the aperture. It may also be noted that form and size constancy were placed in opposition to the texture gradient because the texture was composed of dots of different sizes and shapes (circular and elliptical). Greater curvature was seen with binocular vision for all textures.

The curvature of a semicylinder was also judged in a further study by Smith and Smith (1961), who presented singly and in combination a large number of cues to slant: linear perspective, light and shade, texture gradient, form transformations, monocular movement parallax, binocular disparity, and interposi-

tion. When the edges of the cylinder were obscured, all these cues in combination were not sufficient for veridical judgments. "These results indicate that the gradient theory . . . may be . . . psychologically insufficient for accounting for veridical perceptions of slant in all situations [Smith & Smith, 1961, p. 147]." It is clear that the availability of a set of variables, for example, a combination of binocular disparity, texture gradient, and motion parallax, which uniquely specifies the source, does not insure veridical perception. The fact that perceptions were veridical when the edges of the cylinder were shown may indicate that recognition of a familiar object (a cylinder) was important.

According to Gibson (1951) outline figures produce ambiguous percepts. Because a given retinal shape may be produced by an indefinite number of forms at appropriate orientations, it seems logical that simple retinal form-stimulation should elicit unstable percepts in the absence of other cues. Gibson maintains that texture gradient is necessary for the reliable perception of slant, even though it is not sufficient. Contrary to Gibson's view, experiments by A. H. Smith and his associates indicate that an adequate stimulus for perceived slant is the retinal gradient of outline convergence. Clark, Smith, and Rabe (1955) limited the subject to monocular vision, with motionless head, and required him to judge the slants of outline forms which were presented in succession and at relatively great distance in order to eliminate any cue arising from accommodation. It was found that a trapezoid appeared unslanted and rectangular when slanted so as to project a rectangular retinal image. A trapezoid presented in the frontal-parallel plane (or at 20 degrees) appeared to have the same slant as a rectangle slanted at 20 degrees (or at 40 degrees) so as to project the same retinal image. As physical slant increased, mean perceived slant generally increased, although the mean perceived slants were underestimations (probably due to the absence of other cues for depth). In spite of Gibson's (1951) finding that outline forms are ambiguous, no observer in the experiment under consideration reported that the stimulus forms changed in their apparent shape or slant. The discrepancy in results may be due to the fact that Gibson's figures were drawn on grounds presented in the frontal-parallel plane; Clark, Smith, and Rabe (1955, 1956) set off their outline figures from their backgrounds. With Gibson's figures there would be a conflict between the slant percept produced by an outline figure and the unslanted percept required by the orientation of the ground. There is considerable evidence that the slant of a figure tends to be assimilated to the slant of its ground (Beck & Gibson, 1955; Epstein, Bontrager, & Park, 1962; Metzger, cited in Koffka, 1935, p. 124).

In a further study, Clark, Smith, and Rabe (1956) found that when outline perspective was the only cue, the accuracy of slant perception was greater than when texture gradient was the only cue. Accuracy of perceiving slant did not improve when the gradient of outline deformation was supplemented by adding texture to the stimulus form or to the background or to both. When the ground was textured, perception was less accurate.

Stimulation for the Perception of Depth

Gibson maintains that depth perception does not occur without the perception of a surface (and hence without a gradient of texture, which is required for the perception of a surface). To test this hypothesis, Weinstein (1957) inked out all texture from a set of Gibson's (e.g., 1950b, pp. 184–185) photographs of a large open field with stakes driven

into the ground at various distances from the camera. Judgments of the sizes of the stakes made by the subjects using the impoverished pictures were not significantly poorer than judgments made by the subjects using the normal photographs with a texture gradient present. Using the same type of impoverished pictures, Smith (1958) obtained relatively good size constancy. According to Smith, it would not have been possible for the subject to make size judgments of the distant stake in the impoverished photographs if he had not assumed a level ground-plane common to himself and the test objects. Without such an assumption, the image of an object at the same position in the light pattern could give rise to a percept of a small near object on a hillside or of a larger, more distant object on a level plane. The subjects in Smith's experiment were told to assume that a plowed field extended from their position for a long distance. They were also told that their view would be that of the camera. With these assumptions, and an awareness of the angle of elevation of the eye, they had sufficient information to make size and distance judgments.

Natsoulas (1963) has pointed out that the visual depth experienced in the *Ganzfeld* situation contradicts Gibson's requirement that the visual perception of depth be produced by certain attributes of the textured optical array, especially variations in density, disparity, and motility. Gibson rejects this kind of finding on the ground that, like all experiences taking place in response to impoverished, ambiguous, or equivocal patterns, it is unreliable, varying from observer to observer and from time to time in the same observer. The theory need not explain these experiences because they are a function of something other than stimulation and are not veridical. Natsoulas points out that the experience of depth in the presence of the *Ganzfeld* is not unreliable; it is an unvarying accompaniment of the *Ganzfeld* situation. Gibson, Purdy, and Lawrence (1955) found variability in their subjects' reports because the stimulus field was not entirely homogeneous. Among other things, there was the gross inhomogeneity between the surface of the first sheet of plastic and the optical tunnel seen through the aperture in this sheet. As Cohen (1957) has demonstrated, an inhomogeneity anywhere in the field can reduce the tendency to see a surfaceless fog in other parts of the field. When all inhomogeneities have been eliminated from the field, almost all subjects have reported a surfaceless fog extending away in depth (Cohen, 1957; Hochberg, Triebel, & Seaman, 1951). Not only do experimental results of this kind make it appear that Gibson is wrong with regard to the necessary stimulus conditions for the perception of distance; such findings also contradict the fundamental assumption of this theory that perception, when it is reliable and stimulus determined, is also veridical.

Psychophysical Correspondence under Conditions of Transforming Stimulation

As was the case for static stimulation there are two related questions which need consideration: (*a*) Do transformation sequences yield consistent, relatively unvarying percepts, and (*b*) are these percepts in good correspondence with the objective properties of the source of stimulation? In this section we will consider these questions with regard to motion perspective, perspective transformations, and size transformations.

Motion Perspective

Motion parallax has long been recognized as a depth cue. However, beginning with Helmholtz' (1925, pp. 295–296) well-known description, the emphasis has been on the relative an-

gular velocities of isolated objects at various distances from the subject. This emphasis is reflected in the experimental investigations (e.g., Graham et al., 1948) which typically have used only two objects and two velocities in an otherwise uniform field of view. This approach is also consistent with the traditional "air theory" (Gibson, 1950b, p. 6) of space, as contrasted with the "ground theory" which emphasizes the variations in stimulation which are correlated with a continuous background surface. When the ground or terrain is taken into account there is a continuous flow or gradient of retinal velocities. This gradient involves a continuous transformation of the angular separations between points in the field or, more simply, a transformation of pattern. Gibson uses the term "motion perspective" to distinguish these conditions from those which prevail in the two-velocity case. Motion perspective is presumed to be a stimulus correlate of perceived relative distance in the same way that gradients of optical texture are presumed to be correlates of perceived recession of a surface (Gibson, Olum, & Rosenblatt, 1955). Thus motion perspective is assigned the status of a stimulus for depth. This contrasts with its frequent designation as a "clue" which is effective only when other information makes the assumption of spatial extension reasonable (e.g., Ittelson's, 1960, neo-Helmholtzian analysis).

In order to examine Gibson's interpretation, experiments are needed which isolate motion perspective. Will motion perspective in isolation consistently yield an impression of depth which is compatible with the variability of retinal velocity? An early experiment by Gibson and Carel (1952) produced negative results. The source of stimulation was a battery of scattered lights having a uniform overall density. The battery was in the subject's frontal plane and appeared so when motionless. The lights could be moved across the field with a velocity which decreased from the bottom of the battery to the top. The display was exhibited in the dark, and observation was monocular. The gradient of retinal velocity thus produced should be sufficient to yield the perception of a surface which recedes from the subject from bottom to top. This expectation was not confirmed. A consistent impression of recession, slant, or relative distance was not induced.

Negative results were also obtained by Smith and Smith (1957, 1961) under different conditions. They investigated the conditions necessary for the veridical perception of the convexity of a cylindrical textured surface. Rotation of the cylinder on its horizontal (long) axis was not sufficient to cause the cylinder to appear curved if it appeared flat when motionless.

Subsequent experiments by Gibson, Gibson, and Smith (1959) have also failed to produce unequivocal results. A series of three experiments was conducted with the shadow projector described by Gibson (1957) and Gibson and Gibson (1957). Experiment I was a two-velocity experiment in which subjects were asked to describe what they saw in response to the differential translatory velocity of two circular spots or two superimposed textures composed of elements of indefinite size and contour. Theoretically, the perceived distances of the two spots or textures should be inversely proportional to the optical velocities. The results of the experiment did not confirm this expectation. The magnitude of apparent separation varied greatly between subjects. However, even more damaging was the fact that the direction of the difference in depth was not consistent between subjects, that is, almost 25% of the subjects reported that the spot or surface carrying the lesser velocity appeared nearer. The authors speculate that "perhaps it was

(the) lack of solid continuity or rigid connectedness between the nearer and farther surface which prevented the ideal possibility . . . from being realized [Gibson et al., 1959, p. 46]."

For this reason Experiment II was conducted with a continuous flow of velocity produced by a single randomly textured surface slanted 45 degrees away from the translucent window of the shadow projector which was moved back and forth parallel to the screen. This situation is similar to the one produced by Gibson and Carel (1952). The main result of Experiment II was that most subjects reported a rigid moving plane surface slanting away from them at the top of the field. However, estimates of the slant varied from 12.5 degrees to 60 degrees. Again the magnitude of the apparent separation was highly variable.

Smith and Smith (1963) tested the prediction that the inconsistencies observed by Gibson et al. (1959) would be eliminated by using a combination of two-velocity parallax and a continuous gradient of velocity flow. The results did not support the prediction. The direction of apparent depth agreed with the requirements of the velocity ratio; however, the magnitude of separation in depth of the surfaces was indeterminate. Also, contrary to the results of Gibson et al. (1959, Experiment II) the continuous flow of velocities did not yield the perception of a slanted surface.

The results of the experiments reviewed above do not confirm Gibson's assertion of a strict psychophysical relation between a continuous gradient of optical motion and perceived depth. The experiments failed to provide evidence that motion perspective in isolation is adequate to specify depth, slant, or the relative order of surfaces in depth. It will be recalled that the studies of texture gradients in isolation (Beck, 1960; Gibson, Purdy, & Lawrence, 1955) also yielded inconsistent percepts. Apparently perception is not correlated with individual gradients of stimulation in a 1:1 manner. In light of this observation it seems more promising to study various gradients in combination, for example, motion perspective and textural density, instead of single gradients in isolation. Smith and Smith's (1961) study of the perception of cylindricality illustrates the possibilities of this approach.

Several general remarks are in order concerning the expectation that psychophysical correspondence would be observed when only a single gradient is available.

1. This expectation implicates certain assumptions concerning the properties of the environment. For example, with regard to the expected correlation between apparent recession and optical density, there is the assumption that the elements which constitute the texture of the environment are the same physical size and shape throughout. This is probably a safe assumption insofar as we deal with man-made surfaces. However, when natural terrain is considered there is no compelling reason to believe that this is consistently the case. Many surfaces are admixtures of uniform physical texture and texture composed of elements of variable size and shape. The "ecological validity" (Brunswik, 1956) of the various optical gradients is not known. In fact Gibson and Flock (1962), in a discussion of the illusory closeness of mountains, have offered as an explanation the fact that "in the neighborhood of a mountain, the distant earth-shapes may be much larger than the nearer ones and the usual optical gradient will then be altered [p. 502]." This is presumed to produce the apparent nearness of the mountain despite the contradictory information provided by the concurrent gradients of disparity and motion perspective.

Gibson (1959) feels that the occurrence of illusions and errors in perception will not be explained in the same way as veridical perception. Supplementary theoretical principles are required. As yet the details of this theory have not been advanced. However, a hint of what Gibson's theory would be like with these corollaries added is given by his comments on the apparent distance of mountains (Gibson & Flock, 1962):

> There are many possible types of stimulus-information for distance in a natural optic array, some of them based on constant laws of geometry and optics, others based on probable but variable conditions of geological features, atmosphere, and illumination. Illusory perception will depend on the combination of circumstances which holds for the particular situation, and also probably on the degree to which the attention of the observer has been trained to register the reliable information in the array [p. 503].

This statement sounds very like those made by Brunswik (1956) regarding the learning by the observer of the ecological validity of cues. In fact, it is hard to see how Gibson's system, in its final and complete form, will be able to retain its individual identity.

2. Assumptions of a different sort are implied when other gradients are considered. Two examples will illustrate: (*a*) When the objects under observation are not in the same line of sight, the prediction that differential retinal velocity produced by actual movement in the field will yield veridical apparent relative distance seems to require, in the case of a moving subject, the correct registration by that subject of the relative rates of his movements when viewing the objects whose distances are to be compared. A correct perception of the magnitude of the separation between the objects would involve a correct perception of the absolute rate of movement. This is necessary if the perceptual system is to partial out the various sources of differential retinal displacement. Such correct registration would be a considerable achievement if the subject is undergoing passive locomotion. (*b*) Another example concerns the gradient of relative size. It is evident that this gradient in isolation is ambiguous for the same reason that the visual angle subtended by a single object is ambiguous. The continuous diminution of the angles subtended by a series of objects does not uniquely specify recession since it is equally compatible with a frontal-parallel series of objects which are actually regularly compressed in physical size. The ambiguity is eliminated if the subject assumes that the objects are the same size. Gibson (1950b) seems to recognize this when he writes of the "retinal gradient of size-of-similar objects." Recent experiments by Epstein and Baratz (1964) have shown that the apparent relative distance produced by relative visual angle may indeed be governed by such assumptions.

Continuous Perspective Transformations

The projection of a form or pattern on a non-frontal-parallel surface constitutes a perspective transformation of the form or pattern. In this sense the retinal projection of any form, pattern, or surface which is not in the subject's frontal-parallel plane will constitute a perspective transformation. When either observer or object is in continuous motion the result is a continuous sequence of perspective transformations.

Most studies of the effect of transformation have presented the static end products of transformation, concealing from the subject the antecedent transformation sequence. This set of conditions must represent an atypical case. Behavior in the natural environment is marked by continual shifts of the head and body leading to continuous transformations of the retinal pattern. A predilection for static stimulation also is evident in those few instances where

changing stimulation has been studied. As an illustration, Wallach and O'Connell (1953) in interpreting the kinetic depth effect (KDE) treat the continuous sequence of shadow transformations as a series of independent static projections tied together by memory traces. It will be clear from the earlier discussion that Gibson takes exception to the traditional emphasis on static stimulation. According to Gibson the emphasis should be reversed.

Only a small portion of the evidence on this question has direct Gibsonian origins. Moreover, the implications for Gibson's theory have often not been made explicit by the original investigators. Therefore, their agreement with our conclusions cannot always be assumed. The studies reported by Gibson dealt with slant depth of surfaces. The experiments which derive their impetus from Wallach and O'Connell's (1953) work on the KDE have involved the internal depth and coherence of complex forms, for example, a bent parallelogram. Finally, Langdon's (1951, 1953, 1955) investigations of shape constancy also have implications for Gibson's view.

Slant Depth. Several experiments reviewed in this section used a device called a shadow transformer invented by Gibson and Gibson (1957). A continuous sequence of shadow transformations is produced on a translucent screen by a rotating object placed between the screen and a point source of light which is positioned so as to produce polar projection.

Gibson and Gibson (1957) presented the shadow sequences produced by turning two regular (closed-square, open-square pattern) and two irregular (closed amoeboid, broken amoeboid) forms through 4 degrees of semirotation. There were two main questions: (*a*) Would the sequence of transformations yield the perception of a rigid surface of changing slant? (*b*) Would judgments of the amount of slant correlate positively with the variations in the length of the transformation sequence, that is, the degree of semirotation? The results for all four forms were affirmative on both counts. A separate control group was shown only the motionless end point of a 60-degree transformation. For this group the irregular forms yielded judgments of "no slant," whereas the regular forms were perceived as slanted. However, the amount of slant was greatly underestimated and was considerably less than the slant produced by the 60-degree transformation sequence. Gibson and Gibson concluded that a perspective transformation sequence is sufficient to produce an accurate perception of slant and rigidity regardless of the familiarity or texturedness of the form which carries the motion.

This conclusion cannot be accepted without reservation. The appropriateness of the control condition is questionable. There is little reason to expect that exhibiting the end points should be equivalent to displaying the form in the numerous aspects of its transformation arc. The logical requirement is that no single static aspect or combination of discrete static aspects yield a perception of slant. Therefore, what is required is that a random sample of discrete static transformations be presented. If the subject fails to report changing slant, then we may conclude that a continuous and lawful sequence is essential. The second objection concerns an important difference between the control and experimental conditions which may account for the results. Under the experimental conditions, the subject may safely infer that one and the same object is being presented in various aspects. On the other hand, when isolated static aspects are presented the subject cannot be certain of the continuing identity of the shadow-casting object. The apprehension of physical

identity is crucial since without it the shadow transformations could just as well be perceived as continuous deformations in the plane of the screen.

That these arguments need to be taken seriously is indicated by the results of Sidorsky (1958), White (1962), and Epstein and Mountford (1963). Sidorsky (1958) found that subjects could accurately judge their own apparent rotation in the medial plane when presented with static perspective transformation of a grid-patterned surface as viewed at different pitch angles. Epstein and Mountford (1963) presented the static end products of various amounts of transformation of rectangular representational and nonrepresentational forms. Judged slant was found to vary systematically with the degree of transformation. A general rule may be that any deformation of contour or surface texture which represents a deviation from modal regularity will produce a depth displacement.

White's (1962) experiment is especially interesting. White puts the question this way:

> If the moving presentation is effective merely because it provides more looks at the pattern which are not completely redundant, then the order in which these looks are presented should not make any difference in the accuracy of judgment. If, however, the *transitions* between looks are important, then disrupting the order in which the looks are presented should impair the accuracy of the judgment [p. 75, italics added].

In order to examine this question White showed subjects a motion picture which displayed either an orderly sequence of transformations of an unfamiliar 3D rigid form tumbling about a fixed center, or a scrambled series in which the same frames were randomly ordered. At the conclusion of the movie the subject was asked to select a still picture of the standard form in a new orientation from a set of four similar forms. There was no difference between the judgments based on the ordered and scrambled series. It seems that insofar as accuracy of identification is concerned the regularity of the sequence is not critical.

The results of these experiments have bearing on two salient aspects of Gibson's view. First they raise questions concerning the evidential basis for Gibson's reiterated insistence that the sequence itself is a stimulus, for example, "the *form* of the change of form of the stimulus is what determines the perception [Gibson & Gibson, 1957, p. 137]." Second, doubts arise concerning the criterial validity of the regular perspective transformation sequence. The fact that a source may be mathematically specified by a given sequence in an unequivocal manner cannot be taken to imply that the subject uses this relationship.

Internal Depth. The experimental investigations of the perception of internal depth through motion have been reviewed by Metzger (1953, Ch. 13) and Braunstein (1962).[2] The following discussion will focus on the KDE. Under certain conditions, a sequence of transforming shadows will appear as a rigid form, having internal depth and turning in depth. This has been labeled KDE by Wallach and O'Connell (1953). The above definition is incomplete since it does not specify two important conditions for demonstrating KDE. (*a*) The familiar cues for relative distance must be absent. In contrast with Gibson and Gibson's (1957) situation, this requirement implies the use of parallel rather than polar projection. (*b*) None of the static aspects of the transformation sequence should yield a

[2] Of historical interest is Miles' (1931) discussion of the movement interpretations of the silhouette of a revolving two-blade electric fan. Miles traces the irregular history of this problem back to a description by Nicholson (1802) of a theater exhibition in London.

perception of depth. None of the better-known studies of KDE have provided satisfactory evidence that the second requirement was fulfilled. There is good reason to doubt whether this criterion can be met with figures which have many interior angles and lines. In fact, Fried (1960) tested Wallach and O'Connell's (1953) figures, and concluded that only the single bent rod appeared flat in its various static transformations.

The basic fact of KDE is plainly consistent with Gibson's views. A continuous sequence of transformations is shown to be sufficient to determine the perception of a rigid turning form. However, Gibson's theory demands more than this. First, it requires that variability of perception be small. Second, that the perceived form be in good correspondence with the actual shadow-casting form. Finally, perceived turning should also be in good agreement with the actual turning. Data on these questions can be found in the work of Wallach and O'Connell (1953, pp. 211–212), White and Mueser (1960), Green (1961), and Epstein (in press).

Wallach and O'Connell (1953) asked 20 subjects to make judgments of the form of the shadow-casting object immediately following the kinetic presentation. For the wire helix, subjects were instructed to fashion a piece of wire into the shape of the shadow-casting figure. The reproductions were better than those made by the same subjects while they looked directly at the turning figure. For the 110-degree parallelogram, subjects were asked to select a match from a set of four alternatives which included a copy of the standard and three variants. Seventeen of 30 subjects chose the exact copy as their match. It would appear, therefore, that the perception of internal depth based solely on kinetic stimulation is accurate.

This conclusion was confirmed and extended in an experiment by Epstein (in press). Judgments of internal depth and amount of turning were found to be in good agreement with the actual oscillation and objective depth of a 120-degree bent parallelogram. Accuracy did not vary significantly as the length of the transformation sequence varied from 15 to 85 degrees. Apparently, a sequence produced by oscillation through an arc of 15 degrees is sufficiently informative, and lengthening the transformation sequence contributes nothing to accuracy. It remains to be determined whether this is true independently of the locus of transformation, for example, 40–55 degrees instead of 0–15 degrees. There is the possibility of an interaction of length and locus. Also of interest would be to determine whether identical optical transformations have equivalent perceptual effects regardless of their source, for example, passive movement of the subject compared with movement of the object.

The experiments of Wallach, O'Connell, and Epstein have utilized as their shadow-casting objects figures composed of connected lines varying in length and orientation. However, KDE may also be obtained by rotating a display which consists of an arrangement of unconnected straight lines. For example, four vertical pegs arranged at the corners of a square will appear as a rigid spatial arrangement turning about its center. White and Mueser (1960) studied the subject's ability to reconstruct the arrangement of pegs under variations of the shape and number of pegs constituting the display, the speed of rotation, and the exposure duration. Under all conditions, excepting the static presentations, subjects reported immediate perceptions of a stable rigid arrangement rotating in depth. However, most subjects were unable to reproduce the arrangement with greater than chance accuracy. Additional evi-

dence that accuracy is low for unconnected arrangements may be inferred from the results of Green's (1961) Experiment VI.

It appears that accuracy is enhanced by two properties of the display: complexity, for example, interior angles, and connectedness. Should an explanation of this observation be forthcoming it will probably be intimately related to the answer to a more basic question. Why does the sequence of shadows appear as it does, instead of appearing as a continually deforming two-dimensional pattern? It is not sufficient for Gibson to reiterate that sequences are stimuli for depth and solidity. This may (or may not) be an empirical fact; however, it cannot be adduced as the explanation of the specificity of perception. Nor is it sufficient to demonstrate that the subject can discriminate between two similar transformation sequences, one produced by turning a rigid surface and the other by an elastic surface which has been made to contract (Fieandt & Gibson, 1959).

The answer appears to rest in Gibson's implicit adherence to a variant of the Gestalt principle of *Pragnanz*. According to this principle the 3D rigid appearance is a reflection of the subject's preference for the simpler percept. Or in the alternative terminology of information theory: "Other things equal, that perceptual response to a stimulus will be obtained which requires the least amount of information to specify [Hochberg, 1957, p. 83]." Gibson is cited by Hochberg (1957, p. 83) as one of those who subscribed to this principle. If we are correct in ascribing the minimum principle to Gibson, then this is certainly an instance where the psychophysical hypothesis has been discarded. Perception no longer is strictly a function of stimulation. Rather it is mediated by a selective principle built into the perceptual system. The inclusion of a selective principle robs Gibson's thesis of its attractive simplicity. It also carries the implication that a sequence of transformations which requires less information to specify as a continually deforming two-dimensional pattern will be perceived as such, instead of as a turning rigid form. It is obvious that this outcome would be incompatible with the unamended Gibsonian interpretation. Some evidence that two-dimensional deformations are perceived under these circumstances has, in fact, been reported by Metzger (1959).

The Perceptual Constancy of Turning Shapes. The question of shape constancy has been discussed in great detail elsewhere (Epstein & Park, 1963). Here we will mention only Langdon's studies as they bear on the relationship between transforming stimulation and perception. In normal observation the contour of an object is not given as an instance of static nonchange, or as a series of discrete transformations. Instead the contour is usually represented by a continuous family of perspective transformations. While this is true for the numerous states of the same contour, it need not be true for the differences between the contour shapes of different objects, or the differences between the contours of objects which are actually undergoing physical change. One implication of this analysis is that the traditional formulation of the constancy problem is misleading since it focuses on the relation between discrete static transformations and perceptual stability. A second implication is that continuous transformation should enhance shape constancy. Although Langdon has not followed the route we have just taken, he has conducted a series of experiments which examine the second implication.

For the present analysis the main finding of Langdon's (1951, 1953,

1955) experiments is that the rotation of shapes, under conditions such that a stationary presentation yields no constancy, will be sufficient to restore constancy. The degree of perceptual constancy was positively correlated with speed of rotation or rate of change of shape. These results were obtained both with simple outline figures, for example, a wire circle, and complex "solids," for example, a wire cube. Langdon criticizes the kind of explanation which Gibson might advance and offers an alternative. In any event, Langdon's results provide evidence for Gibson's view that the stimulation most useful for achieving a stable visual world is not constant nonchanging stimulation, but is instead stimulation undergoing continuous change.

Size Transformations

Continuous displacement of a frontal-parallel surface in a radial line from the subject yields a size transformation of the retinal image. Diminution of distance produces a symmetrical expansion of the image, and increasing distance results in a symmetrical contraction. If the surface moves back and forth, then one transformation sequence is a sequential inversion of the other.

While size transformations have long been recognized as cues for relative distance there has been little experimental work on this topic. Some of the early work (e.g., Bourdon, 1902) has been reviewed by Woodworth (1938, Ch. 26) and Boring (1942, Ch. 8). We have uncovered only three recent studies addressed to this question (Ittelson, 1951; Schiff, Caviness, & Gibson, 1962; Smith, 1951). There is some disagreement about whether the perceived radial motion produced by size transformation is regulated by the subject's assumptions about the target. Ittelson (1951) provides evidence that assumptions about, and prior experiences with, the transforming target are important determinants. Smith's (1951) experiment, on the other hand, failed to support this view. However, both of these investigators do provide clear-cut evidence that under appropriate conditions size transformations will yield perceived radial motion. In addition, the studies concerning prediction of time of collision (Carel, 1961) have shown that this effect is lawfully related to the amount of transformation.

The recent experiment by Schiff, Caviness, and Gibson (1962) illustrates the experimental investigation of this cue. Twenty-three monkeys were observed responding to four stimuli: expansion of a circle projected on a screen, contraction of the circle, darkening and lightening of the screen. Two very different modes of behavior were observed as responses to the size transformation. The expanding circle led to alarmed withdrawal, as if to avoid collision. The contracting circle led to exploratory responses. The results for the brightness variations showed that this variable cannot account for the effects of size transformation. Schiff et al. (1962) conclude that size transformations are sufficient stimuli for radial motion. They do not consider the possibility that size transformations merely produced changes in perceived size without affecting perceived distance, and that sudden, rapid changes in perceived size produce withdrawal or approach responses.

Conclusion

In concluding we wish to consider briefly several general questions.

1. The possibility that the constancy hypothesis is sufficient. There were two steps required for the successful resurrection of the constancy hypothesis as an account of veridical perception. First it must be proved that some variable of

stimulation unequivocally specifies the distal source. Secondly it must be shown that the total variance in perception can be attributed to variance in the proximal stimulation.

While Gibson has offered persuasive arguments concerning the specificity of stimulation he has only rarely (e.g., Gibson, Olum, & Rosenblatt, 1955) attempted the mathematical analyses which are demanded. Several unpublished theses by Gibson's students (Hay, 1961; Purdy, 1958) have made admirable efforts to achieve this objective, and they deserve careful study. However, certain physical properties of the environment (such as rigidity) have been accepted as primitive and hence have escaped analysis. What makes this procedure unsatisfactory is not the implausibility of the assumptions. The assumptions are not extravagant, nor is it by itself improper to postulate certain premises in writing geometrical relationships. Our discomfort arises from the fact that some of the properties which are assumed are the very ones which Gibson's theory is intended to explain.

Concerning the psychophysical hypothesis it can be said that Gibson has not proved his case. The experimental data simply do not support the hypothesis of perfect psychophysical correspondence. Nor does the evidence support the contention that perception is "in contact with the environment," that is, veridical, in cases of psychophysical correspondence.

The question of veridicality has a long history in philosophical analysis, and we would wish to avoid the question. However, an evaluation of Gibson's thinking would be incomplete without a discussion of this matter. We are uncertain about the value of introducing the conventionally defined relation of veridicality into a theory of perception. It is clear in Gibson's case that the account of perception is contained in the psychophysical hypothesis which relates percepts to proximal stimulation, and need not involve any assumptions concerning the distal-proximal or distal-perceptual relationships. The introduction of veridicality considerations leads Gibson to distinguish between perceptual events on bases other than the subject's response, or an analysis of stimulation. For example, suppose Subject X looks into an optical tunnel and reports that he sees a solid tunnel. Subject Y is shown a real tunnel which provides identical stimulation, and he too reports that he sees a tunnel. Applying the label "veridical" to Y's report and "nonveridical" to X's report hardly seems a useful exercise.

2. Is a correlational theory sufficient? In reviewing Gibson's *The Perception of the Visual World* Boring (1951) advanced the following criticism:

> What Gibson calls a "theory" is thus only a description of a correlation, a theory which tells *how* but skimps on *why* . . . eventually science must go deeper into the means of correlation, must show in physiology *why* a gradient of texture produces a perceived depth, not merely that it does [p. 362].

Prentice (1951) in his review has made the same criticism of Gibson.

Putting aside the question of whether Gibson needs more physiology or more psychology we would agree that he needs more theory. In our view Gibson has tried to capture too much empirical terrain with too little theory. The consequence has been the unintended development of implicit theoretical propositions which are applied when the explicit theory is confronted by an emergency. These hidden theoretical principles can be employed only at the peril of internal inconsistency. Gibson's theory needs deliberate revision.

REFERENCES

Beck, J. Texture-gradients and judgments of slant and recession. *Amer. J. Psychol.*, 1960, 73, 411-416.

Beck, J., & Gibson, J. J. The relation of apparent shape to apparent slant in the perception of objects. *J. exp. Psychol.*, 1955, **50**, 125–133.

Boring, E. G. *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century, 1942.

Boring, E. G. Review of J. J. Gibson, *The perception of the visual world*. *Psychol. Bull.*, 1951, **48**, 360–363.

Bourdon, B. *La perception visuelle de l'espace*. Paris: Reinwald, 1902.

Braunstein, M. L. The perception of depth through motion. *Psychol. Bull.*, 1962, **59**, 422–433.

Brunswik, E. *Perception and the representative design of psychological experiments*. Berkeley: Univer. California Press, 1956.

Carel, W. *Summary 1961*. (General Electric Advanced Electronics Center Publ. No. R6ELC60) Ithaca: Cornell University, 1961.

Clark, W. C., Smith, A. H., & Rabe, A. Retinal gradients of outline as a stimulus for slant. *Canad. J. Psychol.*, 1955, **9**, 247–253.

Clark, W. C., Smith, A. H., & Rabe, A. The interaction of surface texture, outline gradient, and ground in the perception of slant. *Canad. J. Psychol.*, 1956, **10**, 1–8.

Cohen, W. Spatial and textural characteristics of the *Ganzfeld*. *Amer. J. Psychol.*, 1957, **70**, 403–410.

Epstein, W. Perceptual invariance in the kinetic depth effect. *Amer. J. Psychol.*, in press.

Epstein W., & Baratz, S. Relative size in isolation as a stimulus for relative perceived distance. *J. exp. Psychol.*, 1964, **67**, 507–513.

Epstein W., Bontrager, H., & Park, J. The induction of nonveridical slant and the perception of shape. *J. exp. Psychol.*, 1962, **63**, 472–479.

Epstein W., & Mountford, D. Judgment of slant in response to an isolated gradient of stimulation. *Percept. mot. Skills*, 1963, **16**, 733–737.

Epstein, W., & Park, J. N. Shape constancy: Functional relationships and theoretical formulations. *Psychol. Bull.*, 1963, **60**, 265–288.

Epstein, W., Park, J., & Casey, A. The current status of the size-distance hypotheses. *Psychol. Bull.*, 1961, **58**, 491–514.

Fieandt, K. v., & Gibson, J. J. The sensitivity of the eye to two kinds of continuous transformations of a shadow-pattern. *J. exp. Psychol.*, 1959, **57**, 344–347.

Fried, C. Studies on the kinetic depth effect as a means for presenting three-dimensional information: I. Methodology and selection of forms for study. *USA Ordn. Hum. Engng. Lab. tech. Memo.*, 1960, No. 2-60.

Gibson, Eleanor J., Gibson, J. J., & Smith, O. W. Motion parallax as a determinant of perceived depth. *J. exp. Psychol.*, 1959, **58**, 40–51.

Gibson, J. J. A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, 1941, **38**, 781–817.

Gibson, J. J. The perception of visual surfaces. *Amer. J. Psychol.*, 1950, **63**, 367–384. (a)

Gibson, J. J. *The perception of the visual world*. Boston: Houghton Mifflin, 1950. (b)

Gibson, J. J. What is form? *Psychol. Rev.*, 1951, **58**, 403–411.

Gibson, J. J. Optical motions and transformations as stimuli for visual perception. *Psychol. Rev.*, 1957, **64**, 288–295.

Gibson, J. J. Perception as a function of stimulation. In S. Koch (Ed.), *Psychology: A study of a science*. New York: McGraw-Hill, 1959. Pp. 457–501.

Gibson, J. J. The concept of the stimulus in psychology. *Amer. Psychologist*, 1960, **16**, 694–703.

Gibson, J. J. The useful dimensions of sensitivity. *Amer. Psychologist*, 1963, **18**, 1–15.

Gibson, J. J., & Carel, W. Does motion perspective independently produce the impression of a receding surface? *J. exp. Psychol.*, 1952, **44**, 16–18.

Gibson, J. J., & Cornsweet, J. The perceived slant of visual surfaces: Optical and geographical. *J. exp. Psychol.*, 1952, **44**, 11–15.

Gibson, J. J., & Dibble, F. N. Exploratory experiments on the stimulus conditions for the perception of visual surface. *J. exp. Psychol.*, 1952, **43**, 414–419.

Gibson, J. J., & Flock, H. The apparent distance of mountains. *Amer. J. Psychol.*, 1962, **75**, 501–502.

Gibson, J. J., & Gibson, E. J. Continuous perspective transformations and the perception of rigid motion. *J. exp. Psychol.*, 1957, **54**, 129–138.

Gibson, J. J., Olum, P., & Rosenblatt, F. Parallax and perspective during aircraft landings. *Amer. J. Psychol.*, 1955, **68**, 372–385.

Gibson, J. J., Purdy, J., & Lawrence, L. A method of controlling stimulation for the study of space perception: The optical tunnel. *J. exp. Psychol.*, 1955, **50**, 1–14.

GIBSON, J. J., & WADDELL, D. Homogeneous retinal stimulation and visual perception. *Amer. J. Psychol.*, 1952, **65**, 263–270.

GRAHAM, C. H., BAKER, K. E., HECHT, M., & LLOYD, V. V. Factors influencing thresholds for monocular movement parallax. *J. exp. Psychol.*, 1948, **38**, 205–223.

GREEN, B. F., JR. Figure coherence in the kinetic depth effect. *J. exp. Psychol.*, 1961, **62**, 272–282.

GRUBER, H. E., & CLARK, W. C. Perception of slanted surfaces. *Percept. mot. Skills*, 1956, **16**, 97–106.

HAY, J. C. The importance of optical motions for visual space perception in man. Unpublished doctoral dissertation, Cornell University, 1961.

HELMHOLTZ, H. Physiological optics. In, *Handbuch der Physiologischen Optik*. Vol. 3. (3rd ed. trans. by J. P. C. Southall) New York: Optical Society of America, 1925.

HOCHBERG, J. E. Effects of the Gestalt revolution: The Cornell symposium on perception. *Psychol. Rev.*, 1957, **64**, 73–84.

HOCHBERG, J. E., TRIEBEL, W., & SEAMAN, G. Color adaptation under conditions of homogeneous visual stimulation (*Ganzfeld*). *J. exp. Psychol.*, 1951, **41**, 153–159.

ITTELSON, W. H. Size as a cue to distance: Radial motion. *Amer. J. Psychol.*, 1951, **64**, 188–292.

ITTELSON, W. H. Visual space perception. New York: Springer, 1960.

KOFFKA, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.

LANGDON, J. The perception of changing shape. *Quart. J. exp. Psychol.*, 1951, **3**, 157–165.

LANGDON, J. Further studies in the perception of changing shape. *Quart. J. exp. Psychol.*, 1953, **5**, 89–107.

LANGDON, J. The role of spatial stimuli in the perception of shape. *Quart. J. exp. Psychol.*, 1955, **7**, 19–36.

METZGER, W. Optische Untersuchungen am Ganzfeld: II. Zur Phanomenologie des homogen Ganzfelds. *Psychol. Forsch.*, 1930, 13, 6–29.

METZGER, W. *Gesetze des Sehens*. Frankfurt am Main: Kramer, 1953.

METZGER, W. Was ist wirklich Neu in der Theorie des raumlichen Sehens? *Proc. XV Int. Congr. Psychol., Amsterdam*, 1959, 258–261.

MILES, W. R. Movement interpretations of the silhouette of a revolving fan. *Amer. J. Psychol.*, 1931, **43**, 392–405.

NATSOULAS, T. On homogeneous retinal stimulation and the perception of depth. *Psychol. Bull.*, 1963, **60**, 385–390.

NICHOLSON, W. Narrative and explanation of the appearance of phantoms and other figures in the exhibition of the phantasmagoria: With remarks on the philosophical use of common occurrences. *J. nat. Phil. Arts*, 1802, **1**, 147–151.

PRENTICE, W. C. H. Review of J. J. Gibson, *The perception of the visual world. Amer. J. Psychol.*, 1951, **64**, 622–625.

PURDY, W. C. The hypothesis of psychophysical correspondence in space perception. Unpublished doctoral dissertation, Cornell University, 1958.

SCHIFF, W., CAVINESS, J. A., & GIBSON, J. J. Persistent fear responses in rhesus monkeys to the optical stimulus of "looming." *Science*, 1962, **136**, 982–983.

SIDORSKY, R. C. Absolute judgments of static perspective transformations. *J. exp. Psychol.*, 1958, **56**, 380–384.

SMITH, O. W. Judgments of size and distance in photographs. *Amer. J. Psychol.*, 1958, **71**, 529–538.

SMITH, O. W., & SMITH, P. C. Interaction of the effects of cues involved in judgments of curvature. *Amer. J. Psychol.*, 1957, **70**, 361–375.

SMITH, O. W., & SMITH, P. C. On motion parallax and perceived depth. *J. exp. Psychol.*, 1963, **65**, 107–108.

SMITH, P. C., & SMITH, O. W. Veridical perceptions of cylindricality: A problem of depth discrimination and object identification. *J. exp. Psychol.*, 1961, **62**, 145–152.

SMITH, W. M. Sensitivity to apparent movement in depth as a function of "property of movement." *J. exp. Psychol.*, 1951, **42**, 143–152.

WALLACH, H., & O'CONNELL, D. N. The kinetic depth effect. *J. exp. Psychol.*, 1953, **45**, 205–217.

WEINSTEIN, S. The perception of depth in the absence of texture gradient. *Amer. J. Psychol.*, 1957, **70**, 611–615.

WHITE, B. W. Effect of temporal ordering on visual recognition. *Percept. mot. Skills*, 1962, **15**, 75–81.

WHITE, B. W., & MUESER, G. E. Accuracy in reconstructing the arrangement of elements generating kinetic depth displays. *J. exp. Psychol.*, 1960, **60**, 1–11.

WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

(Received October 25, 1963)

Psychological Bulletin
1964, Vol. 62, No. 3, 197–204

MASOCHISM:

AN EMPIRICAL ANALYSIS

JACK SANDLER [1]

Veterans Administration Hospital, Coral Gables, Florida

Masochistic behavior is defined as a change in response to punishing stimuli which may take the form of simple desensitization at one extreme to continued, self-aversive stimulation at the other. Experimental findings reflecting such changes are cited and some of the conditions necessary for demonstrating these effects are analyzed. The manner in which the scheduling of events may be involved is also suggested. The approach which is advocated seems to offer a fruitful alternative to current popular conceptions of masochism.

The customarily held opinion that man seeks pleasure and avoids pain has been central to most attempts to predict and control human behavior. Recognizing, however, that few situations are either completely punishing or completely pleasurable, learning theorists (Guthrie, 1935; Mowrer, 1960; Skinner, 1953) have expanded this principle to encompass a wider range of human events. In this connection, one argument which has a high degree of consensual validity maintains that some degree of pain is involved in all behavior, even though this may be merely in the form of effort, apprehension, etc. (Brown, 1955). Restating the law of effect then, other things being equal, those situations which involve relatively great pleasure and relatively little punishment will lead to continued functioning, and those events where the reverse is true will result in decreased responding. As Mowrer (1960) states, "It is only when the pain is relatively great and the reward obscure that continued functioning gives rise to confusion [p. 436]." Behavior under these conditions might be termed masochistic.

[1] The author wishes to gratefully acknowledge the assistance and cooperation of Herbert Kimmel and Judson S. Brown in the preparation of this paper.

The concept is no longer restricted to perverse erotic experiences (Reik, 1957) but is often used to describe any situation which encompasses continued functioning in the face of punishing consequences, where common sense dictates a contradiction to a simple punishment effect.

As is true with most clinical concepts, then, the application of the term masochism implies the aberrant, unusual, or bizarre. Most of the current information relating to the topic stems from an accumulation of clinical experiences often formulated within the psychoanalytic framework which emphasizes the importance of guilt, superego conflict, etc. The view advocated here, however, is that masochistic behavior represents an extension of learning principles and a systematic analysis of the phenomenon will reveal the variables which mediate such events. It is the purpose of this paper to provide an empirical definition of masochism which also reflects the predominant features of the clinical event, to review some of the relevant experimental findings which have reported the development of masochistic behavior in the laboratory situation, and to summarize the variables reflected in these investiga-

tions which seem to mediate the phenomenon in question.

A Definition of Masochism

The masochist seems to be the peculiar individual who continues functioning where the punishment is prepotent; that is, where common sense tells us the typical reaction to a given stimulus is avoidance. Any empirical definition must attempt to include this quality of the concept. The crux of the issue rests on demonstrating the typically noxious effects of a stimulus, on the one hand, and the absence or modification of this reaction in the masochistic situation, that is, a change in reaction to what is typically regarded as a punishing stimulus. For our purposes, then, machochism will be regarded as having been demonstrated in those instances where an empirically determined noxious stimulus, defined in terms of a pain response (avoidance, withdrawal, howling, leaping, etc.) does not reveal such effects in another intact member of the same species. Within this framework, masochism may take the form of simple desensitization at one extreme, to the actual performance of an act which produces the noxious stimulus, at the other extreme.

Relevant Research

There have been few attempts to investigate the phenomenon of masochism as such. However, a considerable number of studies have reported findings relevant to the above definition. In some cases, punishment was shown to have little influence on ongoing activities; in others punishment was shown to have facilitated performance and, in still others, punishment was shown to have actually maintained behavior. An arbitrary distinction of these investigations is made in terms of those revealing the above effects.

Punishment

No Effect on Behavior. Perhaps the earliest finding with regard to unusual reactions to punishment is reported by Pavlov (1960). In these instances, the expected response to severe, painful stimuli (intense electric shocks and skin lacerations) was not observed. Pavlov (1960) described the finding in this fashion: "under the very closest scrutiny, not even the tiniest and most subtle objective phenomenon usually exhibited by animals under the influence of strong, injurious stimuli can be observed in these dogs [p. 30]."

Other investigators, working within the classical conditioning framework (Liddell, 1944; Slutskaya, 1928), reported similar findings in children and in a pig, respectively. In each of these instances the noxious event was introduced in typical conditioning stimulus fashion.

Still others, working within the framework of other methods, have confirmed these observations. For example, Mowrer and Viek (1948) were able to reduce the suppressive effects of shock on eating behavior in rats. In this case, rats trained to run a maze for food were shocked 10 seconds after eating. In one group, the shock was terminated by a vertical leap in the air. A second group of rats was administered a physically identical shock independent of any activity. Far greater eating inhibition took place in the group which had no "control" over shock termination. More recently, Miller (1960) was also able to reduce the typically noxious effects of shock on eating behavior in rats. Control subjects, trained to run to a goal for food and suddenly exposed to strong shock at the goal revealed far greater eating inhibition than did rats exposed to gradually increasing shocks at the goal.

Facilitating Behavior. The view that punishment can serve to facilitate behavior under certain conditions has received considerable support. Muenzinger demonstrated such effects in a series of investigations stretching back to the early thirties. The initial finding (Muenzinger, 1934), that a shock which was twice the threshold value for young rats (Muenzinger & Mize, 1933) would actually facilitate the learning of a light-dark discrimination beyond the influence of reward alone, was confirmed in a number of subsequent investigations (Muenzinger & Baxter, 1957; Muenzinger, Bernstone, & Richards, 1938; Muenzinger, Brown, Crow, & Powloski, 1952; Muenzinger & Powlowski, 1951; Muenzinger & Wood, 1935). Likewise, Drew (1938) and Freeburne and Taylor (1952) found that shock enhanced discrimination learning in rats beyond the effects of reward alone.

Rosenblum and Harlow (1963) report a similar effect in connection with a different species operating in a different procedure. They provided six infant rhesus monkeys with free access to a cloth surrogate mother. Two of the monkeys were exposed to 45 seconds of an air blast while on the surrogate, delivered at irregular periods of time. Preliminary observations had demonstrated such a stimulus to have strong aversive qualities for infant rhesus monkeys (p. 83). The results revealed that the punished monkeys spent a significantly longer period of time with the surrogate than did the nonpunished subjects. The authors (Rosenblum & Harlow, 1963) concluded that "an aversive stimulus can augment responsiveness to the surrogate even though the aversive stimulus is coincident spatially and temporally with surrogate contact [p. 84]."

Maintaining Behavior. Perhaps most convincing are those studies which reveal behavior being maintained by noxious stimuli. These efforts are characterized by resistance to extinction as a consequence of punishment occurring during acquisition as well as during extinction.

Several observations are reported by Masserman (1946) which reveal the development of this form of masochistic behavior. Here, cats were trained to press a switch for food reward. Air blasts of gradually increasing intensity were then added as a second contingency to the switch press response. Previously, the air blast was shown to have aversive qualities (p. 75). Finally the food was withdrawn and the cats continued to work the switch for air blasts alone, thus exposing themselves to punishment.

Brown and his associates have conducted a series of studies clearly demonstrating the manner in which behavior can actually be maintained through punishment in a shock escape situation. The assumption here is that electric shock may be regarded as a primary drive source (Brown, 1961). In the first of these attempts, an unpublished observation cited in Mowrer (1950), a rat was trained to run an alley to escape shock. After training, the shock was administered only in a small area just prior to the escape box. The results indicated that the inclusion of this stimulus led to increased resistance to extinction despite the fact that merely remaining in the starting box would have led to shock avoidance. In a similar procedure, Gwinn (1949) found that rats trained in this fashion would run the alley faster and resist extinction longer (thus also receiving more extinction shocks) than did control rats that were not shocked during extinction. Interestingly enough, during extinction, some of the experimental subjects were

exposed to twice the initial shock used in training but this seemed to have little effect in reducing running rates.

More recently, Brown, Martin, and Morrow (1964) have shown that these effects can be even further enhanced by modifying the procedure. In two experiments, rats were trained to escape shocks in a starting box by running down a 6-foot alley into an uncharged goal box. During "extinction" shock was withdrawn from the starting box, but some groups were exposed to shock in part or all of the alley. Control rats received no shock, providing evidence of the usual extinction process. In the first study, the introduction of alley shock had no effect on interfering with running rates. This procedure was then slightly modified (reducing extinction shock, reducing the number of escape trials, and providing for a more gradual transition from acquisition to extinction) in an attempt to enhance resistance to extinction in the alley-shock groups. The results clearly revealed that rats shocked during extinction under these conditions ran more rapidly and over a longer period of time (thus needlessly exposing themselves to painful experiences) than did the control rats. The authors concluded that the repeated approaches toward noxious stimuli developed by these procedures justify describing the behavior as "masochisticlike."

Azrin and his associates (Azrin, 1959, 1960; Azrin & Holz, 1961; Holz & Azrin, 1961) provide further evidence of the extent to which punishing contingencies can contribute to resistance to extinction. The importance of these efforts for the present analysis rests in the fact that the administration of the noxious stimulus is generated by a specific response under the organism's own control. Furthermore, the procedure involves positive reinforcement rather than two extensities of punishment as in the Brown situation. Although the investigations were primarily designed to determine the effects of punishment on operant responding under various reinforcement schedules, the findings are directly relevant to the topic at hand. Their method involves the pigeon pecking response in the typical Skinner box situation. Various schedules of food reinforcement are delivered during which time the pigeon's behavior is shaped up. Following this, shock of varying intensities and duration is introduced as a second contingency to the pecking response. Included in their results (Azrin, 1960) are data which indicate that pigeons will shock themselves literally hundreds of times in order to receive food on an intermittent basis. The extent to which this behavior is maintained is apparently a function of a number of variables including shock intensity and reinforcement schedule. In general, however, they offer strong evidence of an animal's continued responding for stimulation which it would ordinarily avoid, with little interference of typical reinforcement-schedule effects. In the most definitive study of these events (Holz & Azrin, 1961) evidence is offered which reveals that pigeons increase their response rates as a function of punishment alone. In this study, subjects were run under two different daily conditions, one in which the response contingency was variable-interval food reinforcement and continuous, 60-volt shock; and a second, independent, daily extinction session without the reinforcement-shock contingency. The sessions were randomly arranged to eliminate the possibility of the subjects establishing a procedural discrimination. The results indicated that responding for shock and food was always greater than in the alternate sessions. Furthermore, this behavior was main-

tained even when the reinforcement was withdrawn, revealing that response rate was a function of shock alone. Finally, the introduction of shock alone into the extinction sessions resulted in immediate increases in response rate. Holz and Azrin (1961) summarized their findings in this fashion:

> These experiments demonstrate that a relatively severe punishment can increase responding. . . . This procedure of selectively pairing a stimulus with a reinforcer is the usual procedure for establishing a discrimination. This discriminative property that the punishment acquired produced the apparent anomaly. Indeed, the discriminative property came to exert an even greater effect on responding than did the aversive property [p. 231].

Discussion

The experimental findings cited in the preceding section are offered as experimental analogues to various clinical forms of masochistic behavior. In each case, the action of punishment has revealed a change in the expected response to noxious stimuli. In the most dramatic examples, organisms have been found actually working for punishing results. That such effects can be demonstrated has been suggested by several authorities. Guthrie (1935) asserted that aversive stimuli will lose their power to evoke avoidance responses if they are repeatedly presented when responses incompatible with punishment are prepotent. Skinner (1953) proposes two forms of masochism. In discussing how self-injury may be arranged, he states that individuals might expose themselves to aversive stimulation if, by doing so, they avoid even more aversive consequences. In another instance, aversive stimuli might be paired with the reinforcer which follows a given activity. The end result may be that "the aversive stimulus becomes positively reinforcing in the same process [Skinner, 1953, p. 367]."

Surprisingly, this view is in accord with analytical thinking. Thus, Fenichel (1945) states several conditions under which masochism might occur:

> Certain experiences may have so firmly established the conviction that sexual pleasure must be associated with pain that suffering has become the prerequisite for sexual pleasure [p. 357].

He further adds:

> Masochistic activities may follow the mechanism of "sacrifice"; the price beforehand is made to appease the gods and to make them contented at a relatively small cost. Masochistic activities of this kind are a "lesser evil" [p. 358].

Stripped of his analytical terminology, Fenichel seems to be describing the same arrangement as Skinner. More recently, Kelleher and Gollub (1962) have likewise indicated that:

> under appropriate experimental conditions, a positively or negatively reinforcing stimulus can be established as a positive conditioned reinforcer [p. 590].

The factors governing such events have not been explicitly formulated but some clues are offered by the literature. First, in order for the phenomenon to appear, it seems clear that the punishing event must be introduced only after a response has been established in the usual reinforcing paradigm. Secondly, the punishment must be introduced in such a way that the response is not completely inhibited during acquisition. This is precisely the view offered by Martin (1963) wherein he suggests that such an arrangement will lead to greater resistance to extinction than is the case with reward alone. To the extent that Martin is describing the facilitative or maintenance effects of punishment, he is also discussing the masochistic event as herein defined. The number of studies cited revealing masochistic findings suggests that the

phenomenon is genuine and reproducible. The conditions necessary for a punishing stimulus to lose its noxious effects, or to facilitate learning, or to elicit behavior have been explicated by the various investigators as well as summarized by Kelleher and Gollub (1962) and Martin (1963). Obviously, in any given instance certain limiting factors (frequency of association between reward and punishment, strength of punisher, degree of deprivation, etc.) may serve to enhance these effects. The exact nature of these limiting conditions, however, are amenable to experimental analysis.

Such a view offers a general approach to an experimental attack on masochistic behavior. One other condition deserves to be mentioned which may be particularly crucial in the current analysis. For example, some attempt must be made to provide for the maintenance of punished, nonreinforced behavior over longer periods of time than has currently been demonstrated since this would provide even more convincing data to support the argument. One possibility here stems from research investigating the effect of partial reinforcement schedules. There is little doubt that the appropriate scheduling of events has a powerful influence on response durability (Ferster & Skinner, 1957). Thus, behavior can be maintained over long periods of time in the absence of reinforcement through the use of training under intermittent schedules. This factor, together with the establishment of shock as a discriminative stimulus, seems to be responsible for the high rates of self-aversive stimulation which Azrin and his associates report. The author (Sandler, 1962) has confirmed these findings in five marmoset monkeys. Organisms operating under these conditions reveal continued responding in the face of repeated punishment. If observations of these subjects are limited to those periods of responding during which no reinforcement occurs (often as high as 95% of the time) one might even conclude that these are truly masochistic creatures.

Manipulating reward and punishment schedules during acquisition and extinction might provide further data of an even more convincing nature.

Conclusions

There seems to be little doubt that characteristic reactions to painful stimuli can be modified by experimental techniques. Even more to the point, organisms can actually be made to work for punishment. Further efforts to analyze the variables underlying such phenomena within a learning framework will, no doubt, provide additional knowledge regarding the prediction and control of such forms of behavior. Aside from the clinical implications, the view espoused here has bearing on conceptual issues. The typical use of the term masochism carries with it certain unfortunate consequences. Primarily, it focuses on current activities, emphasizing the bizarre qualities of the behavior, to the neglect of other factors which may be even more important determinants (e.g., the reinforcement, the association between the reinforcement and the punishment, etc.). Furthermore, there is an implied assumption regarding the absolute nature of stimuli; that is, certain stimuli always have certain effects and any deviation constitutes an anomaly. Yet common sense observation recognizes the existence of an infinite number of circumstances in which people readily expose themselves to painful experiences. Even presumed, irreversible patterns of avoidance behavior in relatively simple organisms can be modified quite readily (Schneirla, 1959). Finally, the search for clues un-

derlying masochistic behavior usually involves such terms as guilt feelings, superego conflicts, etc.; none of which have proven very fruitful in providing basic insights into the general phenomenon. Current experimental evidence offers, at least, a promising suggestion that such events can be subsumed within a framework of learning principles without recourse to such concepts. Perhaps, then, psychology would be better served if the term masochism was completely eliminated from our vocabulary. When faced with a situation which represents an inconsistency in expected forms of behavior, extensions of the laws of learning might supply some of the answers, and a comprehensive analysis of an individual's reinforcement history might better serve treatment purposes.

REFERENCES

AZRIN, N. H. Punishment and recovery during fixed-ratio performance. *J. exp. Anal. Behav.*, 1959, **2**, 301–305.

AZRIN, N. H. Effects of punishment intensity during variable-interval reinforcement. *J. exp. Anal. Behav.*, 1960, **3**, 123–142.

AZRIN, N. H., & HOLZ, W. C. Punishment during fixed-interval reinforcement. *J. exp. Anal. Behav.*, 1961, **4**, 343–347.

BROWN, J. S. Pleasure-seeking behavior and the drive-reduction hypothesis. *Psychol. Rev.*, 1955, **62**, 169–179.

BROWN, J. S. *The motivation of behavior.* New York: McGraw-Hill, 1961.

BROWN, J. S., MARTIN, R. C., & MARROW, M. W. Self-punitive behavior in the rat: Facilitative effects of punishment on resistance to extinction. *J. comp. physiol. Psychol.*, 1964, **57**, 127–133.

DREW, G. C. The function of punishment in learning. *J. genet. Psychol.*, 1938, **52**, 257–267.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

FREEBURNE, C. M., & TAYLOR, J. E. Discrimination learning with shock for right and wrong responses in the same subjects. *J. comp. physiol. Psychol.*, 1952, **45**, 264–268.

GUTHRIE, E. R. *The psychology of learning.* New York: Harper, 1935.

GWINN, G. T. The effects of punishment on acts motivated by fear. *J. exp. Psychol.*, 1949, **39**, 260–269.

HOLZ, W. C., & AZRIN, N. H. Discriminative properties of punishment. *J. exp. Anal. Behav.*, 1961, **4**, 225–232.

KELLEHER, R. T., & GOLLUB, L. R. A review of positive conditioned reinforcement. *J. exp. Anal. Behav.*, 1962, **5**(4, Suppl.), 543–597.

LIDDELL, H. S. Animal behavior studies bearing on the problem of pain. *Psychosom. Med.*, 1944, **6**, 261–263.

MARTIN, B. Reward and punishment associated with the same goal response: A factor in the learning of motives. *Psychol. Bull.*, 1963, **60**, 441–451.

MASSERMAN, J. H. *Principles of dynamic psychiatry.* Philadelphia, Pa.: Saunders, 1946.

MILLER, N. E. Learning resistance to pain and fear: Effects of overlearning, exposure, and rewarded exposure in context. *J. exp. Psychol.*, 1960, **60**, 137–145.

MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald Press, 1950.

MOWRER, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

MOWRER, O. H., & VIEK, P. An experimental analogue of fear from a sense of helplessness. *J. abnorm. soc. Psychol.*, 1948, **43**, 193–200.

MUENZINGER, K. F. Motivation in learning: I. Electric shock for correct response in the visual discrimination habit. *J. comp. Psychol.*, 1934, **17**, 267–277.

MUENZINGER, K. F., & BAXTER, L. F. The effect of training to approach versus training to escape from electric shock upon subsequent discrimination learning. *J. comp. Psychol.*, 1957, **50**, 252–257.

MUENZINGER, K. F., BERNSTONE, A. H., & RICHARDS, L. Motivation in learning: VIII. Equivalent amounts of electric shock for right and wrong responses in a visual discrimination habit. *J. comp. Psychol.*, 1938, **26**, 177–186.

MUENZINGER, K. F., BROWN, W. O., CROW, W. J., & POWLOSKI, R. F. Motivation in learning: XI. An analysis of electric shock for correct responses into its avoidance and accelerating components. *J. exp. Psychol.*, 1952, **43**, 115–119.

MUENZINGER, K. F., & MIZE, R. H. The sensitivity of the white rat to electric shock. *J. comp. Psychol.*, 1933, **15**, 139–148.

MUENZINGER, K. F., & POWLOSKI, A. F. Motivation in learning: X. Comparison of electric shock for correct turns in a corrective and non-corrective situation. *J. exp. Psychol.,* 1951, **42**, 118–124.

MUENZINGER, K. F., & WOOD, ALDA. Motivation in learning: IV. The function of punishment as determined by its temporal relation to the act of choice in the visual discrimination habit. *J. comp. Psychol.,* 1935, **20**, 95–106.

PAVLOV, I. P. *Conditioned reflexes.* New York: Dover, 1960.

REIK, T. *Of love and lust.* New York: Farrar, Strauss, & Cudahy, 1957.

ROSENBLUM, L. A., & HARLOW, H. F. Approach-avoidance conflict in the mother-surrogate situation. *Psychol. Rep.,* 1963, **12**, 83–85.

SANDLER, J. Reinforcement combinations and masochistic behavior: A preliminary report. *Psychol. Rep.,* 1962, **11**, 110.

SCHNEIRLA, T. C. An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1959.* Lincoln: Nebraska Univer. Press, 1959.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.

SLUTSKAYA, M. M. Converting defensive into food reflexes in oligophrenics and in normal children. *Zh. Neuropatol.,* 1928, **21**, 195–210. Reviewed in G. H. S. Razran, Conditioned reflexes in children. *Arch. Psychol., N. Y.,* 1933, **23**(148).

(Received December 9, 1963)

Psychological Bulletin
1964, Vol. 62, No. 3, 205–209

IN DEFENSE OF DISSONANCE THEORY: REPLY TO CHAPANIS AND CHAPANIS

IRWIN SILVERMAN

State University of New York at Buffalo

This paper evaluates the critical review of the dissonance literature by Chapanis and Chapanis (1964) and concludes that these authors, for the most part, do not accomplish what they purport to, i.e., the presentation of alternative explanations of the findings they deal with. The "methodological inadequacies" discussed by the Chapanises are viewed as acceptable procedures and necessary components of the inductive process. An alternative explanation for the popularity of dissonance theory is offered and the current status of the model is reconsidered.

For the most part, what Chapanis and Chapanis (1964) offer to substitute for existing interpretations of the data of dissonance research may be described as the first half of an alternative explanation. They supply a novel account of the effect of the experimental manipulation upon the mediating processes of the subject, that is, they ask the question, "how can we be sure that the experimental situation has been successful in creating dissonance and dissonance alone? [p. 3]," but they do not complete the process of explanation by indicating how the intervening variable which they have invoked in lieu of dissonance accounts for the subject's response. Further, in several cases the premise on which they base their reinterpretation of the mediating variable appears highly questionable. Let us consider each of their critiques in turn.

Aronson and Mills (1959) performed an experiment whereby they varied degree of punishment, in the form of embarrassment, that the subject endured in order to participate in what was arranged to be a dull group discussion. Their findings were consistent with the prediction that the high punishment group would tend to more greatly enhance the value of the discussion in order to resolve the greater amount of dissonance between the pain they had suffered and the reward they had obtained. Chapanis and Chapanis (1964) question whether high embarrassment did provide an experience of punishment and maintain that "pleasureable cognitions" may have been introduced, stemming from "a feeling of successful accomplishment in overcoming a painful obstacle [p. 5]." Consider the implications of this line of reasoning. It may be extended to any experimental manipulation labeled as punishment and it insists that these be reevaluated in terms of their rewarding effects, a deduction which is generally inconsistent with the data on escape and avoidance conditioning. Even if the possibility is allowed that there is reward value in the punishment condition, there is no explanation of how this variable determines the subject's response. To this end the Chapanises (1964) leave us solely with the statement: "There is no need to postulate a drive due to dissonance if a *pleasure principle* can account for the results quite successfully [p. 5]." Bear in mind, however, that the criterion variable was not the subject's evaluation of the painful or pleasurable initiation, but the group discussion following. An alternate model would have to demonstrate how the

pleasurable effects associated with the embarrassment condition were transmitted to the contiguous but independent event of the discussion. This may have interesting possibilities for some theoretician, but until these are explored, dissonance theory affords the only complete explanation.

The authors then turn to the original test by Festinger and Carlsmith (1959) of the hypothesis that size of incentive for engaging in attitude-discrepant behavior is negatively related to subsequent attitude change. Here the Chapanises consider that the high incentive ($20.00) was an "implausible reward," and thus "the results fall neatly into the pattern of all previous and more extensive experiments on the effect of credulity on pressures to conformity (Fisher & Lubin, 1958) [p. 7]."

We find, however, that the implausibility which the authors refer to in evaluating the dissonance experiment, that is, suspicion of the experimenter's motivation for giving a $20.00 incentive, is not directly comparable to the credulity variable in the studies cited by Fisher and Lubin (1958). The latter refer to the effects upon conformity of "the size of discrepancy between the subject's pre-influence judgment and the subsequent judgment originating from an influence source [p. 230]," for example, the distance between lines in an Asch-type paradigm. Decreased conformity at the higher levels of discrepancy is attributed to disbelief on the part of the subject that the influence source is giving true judgments. The only way we might link these two sets of results by a credulity explanation is by stating the principle that any suspicion of the experimenter's intentions in an influence study produces nonconformity, a proposition which may be developed into an alternative interpretation of Festinger and Carlsmith's (1959) results, if it were first corroborated.

Rosenberg (1963), in fact, has provided an explanation of the Festinger and Carlsmith findings, based in part on a credulity notion, in which it is assumed that high incentive subjects perceive the actual purpose of the study as an investigation of their capacity to be bribed into opinion change and respond in a socially desirable manner. It should be pointed out, however, that Brehm and Cohen (1962) anticipated a possible explanation of the Festinger and Carlsmith results in terms of suspicion of the $20.00 incentive and replicated their findings using high and low incentives of $1.00 and $.50. More recently, a student of the present author (Lependorf, 1964) obtained significantly less opinion change in a $.50 incentive group compared to a $.05 incentive group.

Further in the article the Chapanises "reinterpret" the findings of Aronson (1961) that subjects expending less effort to obtain both nonrewarding and rewarding objects would tend more to enhance the characteristics of the latter. They attempt to explain these effects by noting that the low-effort group was also a higher rate-of-reward group (though total amount of reward was the same for both). Again there is no attempt to link the intervening variable to the dependent variable, but the implication is that rate of reward is positively related to the magnitude of secondary reinforcement effects. Do the existing data substantiate this notion? The studies by Skinner and his students (cf. Ferster & Skinner, 1957) comparing the effects of continuous and periodic reinforcement conditions suggest that the opposite relationship exists.

In one case the authors do not begin to formulate an alternative explanation. In evaluating Brehm's study (1959) in which boys who had consented to eat a disliked vegetable became more positive about it when told

that their parents would be informed of the event, the Chapanises note simply that "The key to the problem most likely lies in the expectation these boys had about the effect of the letter on their parents and on themselves [p. 8]." This is not inconsistent with the rationale underlying Brehm's hypothesis; however, this author offered an interpretation of the expectation which allowed him to predict the obtained effects. The critical point of the Chapanises' objection appears to be that "the design of the experiment does not allow us to find out what this expectation was [p. 8]." Need it be pointed out that the mediating variable, by definition, does not have to be operationally represented; that it functions in the hypothetico-deductive method to enable the experimenter to predict the dependent variable, which Brehm seems to have accomplished very well.

Chapanis and Chapanis spend a large part of their review on a study performed by Festinger (1957, p. 164) and replicated by Cohen, Brehm, and Latané (1959) demonstrating the relationship between magnitude of induced dissonance and avoidance of further dissonance. The crux of their critique is that Festinger offers several explanations for various aspects of the data which, taken together, could account for results in any direction. An important consideration, however, is ignored by the reviewers.

Festinger's study investigated the behavior of subjects who had chosen the side they believed advantageous in a two-person game based solely on probabilities, and who were either winning or losing. Subjects of the former group were considered to have no dissonance between their choice of side and its consequences. The others were assumed to have dissonance in amounts directly related to the amount of money lost. On the basis of these assumptions, Festinger was able to predict the precise form of the relationship, which approximated a fourth-order parabola, between amount of winnings or losings and time spent studying a graph which purported to show the actual probabilities for winning for either side. Festinger considered, however, that though the time scores for the winners, relative to the losers, were as predicted, they were higher in terms of absolute values than his deductions would lead him to expect. On this basis he proposed that an additional set of dissonant cognitions may have been introduced in this group between their performance and information contained in the graph to the effect that they were on the side with a lower probability of success. Thus, two sets of cognitive elements were assumed to be operative in this study, and each considered alone would predict an opposite pattern of relationships between the independent and dependent variables. The essential point, however, is that the relationship that was observed was hypothesized on the basis of deductions concerning one of these sets of cognitions. Deductions concerning the other were invoked, post facto, to account for an aspect of the larger effects of the manipulation which was not consistent with theoretical expectations. Further, the propositions of Festinger's post-hoc analysis meet the sole criteria for their justification; they lend themselves readily to investigation. A partial replication of the original experiment is called for in which dissonance between performance and information contained in the graph is eliminated by modifications of the latter.

In one instance alone, the Yaryan and Festinger study (1961), the Chapanises provide us with an alternative explanation of the effects of the independent variable upon the dependent variable, and it is agreed that this experiment needs to be redesigned before

it may be considered support for dissonance theory.

In the second section of their review, Chapanis and Chapanis (1964) challenge what they term "methodological inadequacies in the analysis of results" of which they state, "the most important is the rejection of cases [p. 12]." The examples they cite, however, do not contain evidence of data exclusion on the basis of the dependent variable, which would certainly constitute a methodological flaw, but involve, for the most part, the acceptable and widely used procedure of preselecting subjects by some criterion that enables test of the hypothesis. The cases where these criteria are established and applied after inspection of the results for a larger sample represent a different aspect of the scientific method than the instance described above, but inasmuch as these criteria are other than the dependent variable, this may be regarded as induction rather than "methodological inadequacy." In the areas of behavior probed by dissonance theory we are a long way from making predictions with anything approaching one to one precision, and until this time it will be incumbent upon the researcher to try to account for data which do not fit his theory in a manner which will lead to further research. When this process involves the attempt to partial out potential sources of error variance, it will often require observation of the effects of the independent variable upon a more limited subject sample defined on the basis of the proposed error factor. Similar practices may be noted in any area of psychological research where they are equally justified and productive. Consider the various studies where interpretations were based upon the discovery that predicted relations held for one sex and not the other (e.g., Janis, 1954; Sears, Whiting, Nowlis, & Sears, 1953).

A valid objection to subject selection, whether it occurs as part of the deductive or inductive process, requires that the objector demonstrate how the selection procedure may account for the results in a manner other than that proposed by the experimenter. The Chapanises do so in just one case, the study by Brehm and Cohen (1959).

In their conclusion the Chapanises attribute the greater attention given to dissonance theory, as compared to the rather similar models proposed by Heider (1958), Osgood and Tannenbaum (1955), and Newcomb (1953), to the "generally deceptive" simplicity of the former. A more feasible explanation might be offered. All four of these models are based upon the principle of "cognitive balance" as a motivational construct. The latter three systems, however, are limited in their application to imbalance which may occur when two attitude objects, here including other persons or the self as objects, are linked in what Osgood and Tannenbaum term an assertion (unit or sentiment relation in Heider's theory). Festinger's theory accounts for this event but extends the balance principle as well to a multiplicity of apparently diverse areas of behavior. This is accomplished by the artful process of considering the cognitive mediator which is interposed in the person's response to all complex stimuli, including his own behavior.

In one respect we agree that the dissonance model must become less "simple" if it is to continue to function as a useful theoretical instrument. If we accept the statement that the value of a theory resides both in the magnitude of testable deductions generated from it and the extent and diversity of the phenomena that it can explain, then, by these criteria, dissonance theory appears to be a good theory. A good theory, however, by its own defini-

tion, cannot remain intact. Every research based upon it, by virtue of the variance in the data which is not accounted for, should lead the investigator back to the theory to revise and elaborate and thus develop new hypotheses for test. It is this interplay between induction and deduction which contributes to the progress of a science, with theory functioning as an instrument of logic in the service of extending the possibilities for acquiring data.

Many may agree that we have had an abundance of deduction from dissonance theory, that is, hypotheses based directly on Festinger's original principles, but a relative paucity of induction, that is, attempts to reconstruct and extend the theory so that it may incorporate more components of the behavior under scrutiny. For example, the study of factors determining which mode of dissonance reduction will be employed in a given situation remains largely outside the province of our conceptual schemas, though Rosenberg and Abelson (1960) offer a model directed to some aspects of this question. Psychology appears to be ready for neodissonance theory, by whatever name we choose to call it, just as it has progressed to neobehaviorism and neo-Freudianism.

REFERENCES

ARONSON, E. The effect of effort on the attractiveness of rewarded and unrewarded stimuli. *J. abnorm. soc. Psychol.*, 1961, **63**, 375–380.

ARONSON, E., & MILLS, J. The effect of severity of initiation on liking for a group. *J. abnorm. soc. Psychol.*, 1959, **59**, 177–181.

BREHM, J. W. Increasing cognitive dissonance by a *fait accompli*. *J. abnorm. soc. Psychol.*, 1959, **58**, 379–382.

BREHM, J. W., & COHEN, A. R. Re-evaluation of choice alternatives as a function of their number and qualitative similarity. *J. abnorm. soc. Psychol.*, 1959, **58**, 373–378.

BREHM, J. W., & COHEN, A. R. *Explorations in cognitive dissonance.* New York: Wiley, 1962.

CHAPANIS, NATALIA P., & CHAPANIS, A. Cognitive dissonance: Five years later. *Psychol. Bull.*, 1964, **61**, 1–22.

COHEN, A. R., BREHM, J. W., & LATANÉ, B. Choice of strategy and voluntary exposure to information under public and private conditions. *J. Pers.*, 1959, **27**, 63–73.

FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

FESTINGER, L., & CARLSMITH, J. M. Cognitive consequences of forced compliance. *J. abnorm. soc. Psychol.*, 1959, **58**, 203–210.

FISHER, S., & LUBIN, A. Distance as a determinant of influence in a two-person serial interaction situation. *J. abnorm. soc. Psychol.*, 1958, **56**, 230–238.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

JANIS, I. L. Personality correlates of susceptibility to persuasion. *J. Pers.*, 1954, **22**, 504–518.

LEPENDORF, S. The effects of incentive value and expectancy on dissonance resulting from attitude-discrepant behavior and disconfirmation of expectancy. Unpublished doctoral dissertation, State University of New York at Buffalo, 1964.

NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393–404.

OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.

ROSENBERG, M. J. An evaluation of models for attitude change. Paper read at American Psychological Association, Philadelphia, 1963.

ROSENBERG, M. J., & ABELSON, R. P. An analysis of cognitive balancing. In M. J. Rosenberg et al., *Attitude organization and change.* New Haven: Yale Univer. Press, 1960.

SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, P. S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135–234.

YARYAN, RUBY B., & FESTINGER, L. Preparatory action and belief in the probable occurrence of future events. *J. abnorm. soc. Psychol.*, 1961, **63**, 603–606.

(Received April 13, 1964)

Psychological Bulletin
1964, Vol. 62, No. 3, 210–213

CORRECTION FOR ATTENUATION AND THE EQUIVALENCE OF TESTS[1]

MARK W. STEPHENS
Purdue University

AND

DOUGLAS P. CROWNE
Northwestern University

Block (1963) suggested that in estimating the equivalence of tests E should correct for unreliability to give a more appropriate estimate of the "conceptual equivalence" of the tests. In the present paper it is pointed out that: a distinction should be made between conceptual equivalence and functional equivalence; functional equivalence is of prior importance in research seeking relationships of other variables to the variable represented by the tests; in estimating functional equivalence it is not appropriate to correct for unreliability; and, in fact, correction for attenuation may often lead to overgeneralization of results.

Block (1963) suggests that one should correct for attenuation when estimating the "conceptual equivalence" of tests. He proposes that a better estimate of the degree to which two tests measure the same or highly similar concepts is given by correcting for the unreliability of each test. If not corrected for test unreliability, he argues, the "raw" correlation between two tests portrays with unnecessary gloom their convergence on the variable they are presumed to reflect.

The purpose of this note is not to defend the papers (Crowne & Stephens, 1961; Crowne, Stephens, & Kelly, 1961; Wittenborn, 1961) on which Block bases his argument, but to clarify three important issues: the distinction between *functional* and *conceptual* equivalence, the inappropriateness of correction for attenuation when functional equivalence is in question, and the prior importance of functional equivalence.

Functional versus Conceptual Equivalence

The equivalence of different tests of (presumably) the same variable is a critical question in research design and in evaluating and generalizing research results. "Equivalence" can have at least two meanings. The equivalence of primary concern in the Crowne and Stephens (1961) paper was functional equivalence.

Measurement operations are functionally equivalent only to the extent that they are actually interchangeable. Two or more tests are equivalent in this sense when they intercorrelate so highly that whatever residual or noncommon variance remains is of negligible magnitude. More rigorously, functional equivalence is established when one finds (preferably over a series of studies) that two tests are so highly related that any results found with the one will be found with the other.

In brief, the assessment of functional equivalence is wholly empirical and conservative: tests are functionally equivalent to the extent that they have been found to be so; that is, to the extent that scores on one are predictable from scores on the other. To achieve functional equivalence is to assure that one can safely generalize from results with one test to findings with another and combine findings with different tests.

Block suggests another meaning of equivalence—conceptual equivalence. As

[1] The writing of this paper was supported in part by Grant No. G25113 from the National Science Foundation to the junior author.

he defines it, two tests are conceptually equivalent if, "giving due weight to the unreliabilities present, these measures . . . reflect the same underlying dimension [p. 152]." How does one determine that two tests reflect the same underlying dimension? This must, of course, be a matter of inference—and permit alternative inferences. Even a minimal observed relationship could suffice to support the claim of conceptual equivalence—although it would of course represent weak evidence for the claim. In other words, two tests may be conceptually equivalent without even approximating functional equivalence.

In effect, the criterion for functional equivalence is much more rigorous than that for conceptual equivalence—and properly so, since the sole concern of functional equivalence is methodological interchangeability and empirically grounded assurance of the generalizability of results. To infer conceptual equivalence from low to moderate observed relationships can be a step toward adding "faith validity" to "empirical validity."

If more is implied by the notion of conceptual equivalence than simply a less demanding criterion of agreement, then the additional meaning must involve the theoretical relations of the tests to the concept they are presumed to reflect. In this case, one would examine the theoretical derivation of two (or more) tests and, by logical analysis, compare their relations to the construct. If functional and conceptual equivalence can be meaningfully distinguished, the distinction lies here: that conceptual equivalence entails a theoretical analysis while the demonstration of functional equivalence is an empirical question.[2] This is not to derogate the role of theoretical analyses but to keep distinct the empirical and theoretical processes.

Functional Equivalence and Correction for Attenuation

The difference between functional and conceptual equivalence is, in fact, best illustrated in the specific case of correction for attenuation. Suppose, for example, that two tests of self-acceptance (SA_1 and SA_2)—or, more conservatively, of self-evaluative behavior —intercorrelate .50 and that the reliability of each is .71.[3] Correcting for the unreliability of the two tests, the intercorrelation becomes .70. This corrected correlation coefficient would suggest stronger evidence that the tests are "measuring the same thing"; that is, that the behavioral samples represented by the two are drawn from the same population of behaviors and/or that one can meaningfully speak of "self-acceptance" as being measured, or reflected, by each test. Here, correction for unreliability results in a more heartening index of the relationship of two relatively unstable instruments and could be used to support the claim of "conceptual equivalence" of the tests.

However, suppose that then one wants to determine the relationship of self-acceptance to anxiety, and one correlates, say, Manifest Anxiety (*MA*) scale scores with SA_1 scores. The Pearson r turns out to be $-.30$, significant at the .05 level. How likely is it, on the basis of this evidence, that SA_2 scores will be significantly correlated with *MA* scale scores? The chances are considerable (certainly greater than .05 and greater than the corrected r of .70 would suggest) that *MA* scale scores

[2] Jessor and Hammond's (1957) criteria for the adequacy of the theoretical linkage between test and construct would seem to be neatly suited to the problem of establishing conceptual equivalence.

[3] It is unimportant here to specify what kind of "reliability" is meant, although, of course, this may often be a critical question.

and SA_2 scores will not be related.[4] What has been learned is that (*a*) 25% of the actual (observed) variance in scores on SA_1 is common to variance in scores on SA_2 and (*b*) 9% of the variance in SA_1 scores is common to *MA* scale scores. From this evidence alone one can hardly be assured that 9% or more of the variance in *MA* scale scores will be common to SA_2 scores. Such an assumption, however, would have been encouraged by correcting for the unreliability of SA_1 and SA_2.

In computing functional equivalence, it is irrelevant what portion of the 75% of variance in SA_1 not common to SA_2 scores is unreliability, what is specific method variance, or what is associated with a conceptually truly distinct variable. Isolating these separate sources of noncommon variance is, of course, important for further research—for example, in attempts to improve the tests—but it is not relevant in denoting the actual relationships between scores on the tests. The point at issue here is that regardless of the source of disagreement among actual test scores, any disagreement still decreases the empirical basis for assuming that an equivalent relationship will be found between *MA* scale scores and scores on SA_2—on self-acceptance tests in general.[5]

Functional Equivalence and the Interpretation of Results

The purpose of this part of the Crowne and Stephens (1961) paper was to caution against overoptimism in generalizing from results with one test and assuming too quickly that different tests of purportedly the same variable will have equivalent relationships to other variables. This position may seem overcautious. However, the research-time cost of the conceptual cul-de-sac is high enough to recommend caution when generalizing from immediate results. Inappropriate correction for attenuation could easily encourage overconfident generalization across "measures of the same thing."

Block (1963) suggested that "By the straightforward application of established procedures for improving the reliability of measures, we may fully expect these theoretical (corrected) correlations to become actuality [p. 155]." However, it is always quite possible that the anticipated correlations will not be realized. High reliability is not always readily achieved: *vide* the MMPI! A further complication may arise when, in attempting to increase the factorial purity of tests, we actually succeed in reducing the relationship between them. This will in fact happen to the extent that the original between-test relationship was mediated by similar items in each test which were not associated with the primary factor represented by each test. In any case, the point at issue is that these correlations are not assured until they have in fact been obtained.[6] The tests currently available, on which we must rely, are still relatively unreliable and method contaminated. They correlate imperfectly with other measures, even of the "same

[4] It should be noted, however, that there is no way to estimate this probability exactly.

[5] In this context, "correcting" for unreliability can in fact lead to another "attenuation paradox": holding the "validity," or actual (observed) agreement with some criterion, constant, the "conceptual validity" of a test increases as its reliability decreases.

[6] In fact, employing the established procedures for improving (intratest) reliability—namely, selecting and discarding items on the basis of agreement with total test score—will maximize the risk of the "attenuation paradox." It will almost certainly lead to progressive narrowing of the population of items represented by the test with a consequent reduction of the relationship of that test to other tests.

thing," and the meaning of the observed relationships between personality tests is still an open and critical question in personality assessment. To consider tests in terms of their "conceptual equivalence"—that is, to restrict consideration of equivalence to the hypothetical "limiting case of perfect reliability"—may often obscure the very questions which are most important to answer. Conceptual equivalence does not answer the primary question of the actual interchangeability or functional equivalence of measures: It does not directly indicate what relationships can be expected between imperfectly correlated measures of the same thing and an external variable.

Further, referring to conceptual equivalence in this sense can obscure the possibility (often considerable) that method and response set factors may be responsible for much of the observed covariation. In the Crowne, Stephens, and Kelly (1961) study, for example, the consistent ordering of subjects on a variety of self-acceptance tests appeared to be chiefly attributable to the effects of a social-desirability response set.

If one is interested in the relationship of some other variable to self-acceptance, it will be efficient in the long run to assess this relationship independently with several different self-acceptance tests, as Campbell and Fiske (1959) have pointedly argued. If these yield equivalent relationships to the extrinsic variable, one can then of course be confident that this variable is related to self-acceptance—and confident of the functional and conceptual equivalence of the self-acceptance tests in this relationship. Without this multiple-measurement strategy, when the conceptually equivalent tests are in only moderate agreement (uncorrected for attenuation), there is a high risk of overgeneralization and overconfidence in the conceptual significance of the results.

REFERENCES

Block, J. The equivalence of measures and the correction for attenuation. *Psychol. Bull.*, 1963, **60**, 152–156.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.*, 1959, **56**, 81–105.

Crowne, D. P., & Stephens, M. W. Self-acceptance and self-evaluative behavior: A critique of methodology. *Psychol. Bull.*, 1961, **58**, 104–121.

Crowne, D. P., Stephens, M. W., & Kelly, R. The validity and equivalence of tests of self-acceptance. *J. Psychol.*, 1961, **51**, 101–112.

Jessor, R., & Hammond, K. R. Construct validity and the Taylor anxiety scale. *Psychol. Bull.*, 1957, **54**, 161–170.

Wittenborn, J. R. Contributions and current status of *Q* methodology. *Psychol. Bull.*, 1961, **58**, 132–142.

(Received October 14, 1963)

Psychological Bulletin
1964, Vol. 62, No. 3, 214–216

RECOGNIZING ATTENUATION EFFECTS IN THE STRATEGY OF RESEARCH

JACK BLOCK
University of California, Berkeley

In the preceding paper, Stephens and Crowne argue that the investigator should not recognize the influence of unreliability in determining the obtained correlation between nominally related measures. The present note reiterates our earlier suggestion that a balanced, sophisticated view of a research area should pay due attention, among other considerations, to the way the attenuation effect may be determining obtained correlations. Proper application of the correction for attenuation has the heuristic effect of informing the researcher whether a raw correlation can be ascribed to the poor but improvable quality of his measures, or whether definite conceptual differences between the measures being related are likely to be involved.

Stephens and Crowne (1964) elaborate both their own (Crowne & Stephens, 1961) and my orientation (Block, 1963) toward the evaluation of research findings in directions that require some countering comments. I shall be brief because, more than a little, we are reduced to claims and counterclaims.

Their particular definition of the *functional equivalence* of tests in terms of correlational interchangeability is acceptable to me for the present purposes and is in due contrast with my earlier usage of the term, *conceptual equivalence*, as the establishment that "measures . . . reflect the same underlying dimension." As one of the factors to be considered in forming a balanced judgment as to the dimensional equivalence of two measures, I reminded Crowne and Stephens of the presence and inevitable effect of error variance to lower the apparent correspondence between measures. This unweighted consideration, coupled with an incompleteness of data reporting, permitted Crowne and Stephens to construe the measurement situation in the realm of self-acceptance in terms different than I believe to be appropriate.

The present divergence appears to me to involve three main points: Should the criterion of functional or of conceptual equivalence have been employed in the first paper in the series? More generally, is functional equivalence of prior importance to conceptual equivalence, as argued by Stephens and Crowne? Is it irrelevant, as Stephens and Crowne suggest, to identify the various sources of variance underlying a relationship?

I choose not to respond to the mentions by Stephens and Crowne of the "attenuation paradox," social-desirability response set, or their hypothetical example because discussion of these less essential points would involve considerations far removed and inordinately complicating.

1. Stephens and Crowne assert that in their first article, they properly were concerned with functional rather than conceptual equivalence. I assert the opposite. I read their paper then—and now—as an attempt to examine in depth the extent to which various proposed measures of self-acceptance could be viewed as alternative operational manifestations of a common construct. In numerous statements (e.g., their section on Definition of the Construct, and

elsewhere), Crowne and Stephens explicitly worry the nature of the concepts being studied, as these are indexed by divergent operations. In their methodological critique, Crowne and Stephens (1961) announce their focus as on "basic logical and psychometric considerations [p. 106]," suggesting an intention to identify and evaluate the several factors influencing particular summary coefficients. Such sustained mention of the problems of supporting the relevance of measures to a construct can only mean, I submit, preoccupation with conceptual or dimensional equivalence. The sufficiently involved reader may wish to resort to the article in dispute and form his own judgment on this point of difference. If he concludes, with me, that the concerns of Crowne and Stephens, in fact, were with regard to conceptual equivalence, then it follows their usage of functional equivalence as an evaluative criterion was narrow and limiting since the influence of error variance was not recognized.

In workaday, applied situations, such as the screening of masses of army recruits, I can see where the notion of functional equivalence has its place and its justification. But not when we are seeking to improve the conceptual basis of our science and organize a welter of findings.

2. Stephens and Crowne support their claim that functional equivalence has a prior importance in research evaluation essentially by value judgment. Thus, where functional equivalence is assigned the virtues of being "wholly empirical," "conservative," and "rigorous," the effort to seek conceptual equivalence is said to be a matter of "alternative inferences," "faith validity," and "weak evidence." [1]

Whatever banner one flies—whether one insists on functional equivalence or seeks conceptual congruence—ambiguous decision problems requiring best guesses confront the investigator or evaluator. These decisions are made on inferential grounds involving considerations going beyond the specific data situation. Inevitably, choice entails costs as well as gains. Specifically, when Stephens and Crowne query: "How does one determine that two tests reflect the same underlying dimension?" they are bothered by the nonexistence of an absolute rule for deciding yes or no. This unhappiness with ambiguity stems from their recognition that, in fact, a continuum of likelihoods underlies any dichotomous decision reached. But, by the same token, what degree of functional equivalence shall we require before deciding that functional equivalence has been achieved? Stephens and Crowne do not say but the parallel exists and may not be compartmentalized out of consideration. When Stephens and Crowne remark that "even a minimal observed relationship could suffice to support the claim of conceptual equivalence," it is easy to paraphrase a rejoinder: even a minimal observed difference could suffice to support the claim of functional nonequivalence.

Stephens and Crowne continue to caution against the cul-de-sac of presuming equivalences prematurely. I commend caution but not overcaution. Simply for balance and perspective, the cul-de-sac of dwelling on differences that are really inconsequential must be recognized. There are dangers of un-

[1] In the same helpful vein, it may be replied that the conceptual evaluation approach is integrative, dedicated toward theory building, and separates the empirical wheat from the empirical chaff. Adherence to functional equivalence, on the other hand, is rigidly antitheoretical, a reversion to naive operationism, and readily confused by the noise in the system.

criticality and abuse in leveling differences to unimportance. The discernment of congruence should not be achieved by projective means. But if the goal of convergent and discriminant validification via multiple-measurement methods is agreed upon, then I would expect this goal to be achieved most rapidly if the contribution of attenuation effects to the analytical picture are soberly considered rather than totally disregarded.

3. I am alarmed when Stephens and Crowne now assert that, with an emphasis on functional equivalence, "it is irrelevant what portion of the . . . variance . . . is unreliability, what is specific method variance, or what is associated with a conceptually truly distinct variable." I believe most psychologists would find the distinction among these three sources of variance of paramount importance. Entire strategies of subsequent research may hinge upon a judgment or knowledge of the size of variance components. With regard to the correction for attenuation, it is useful in that it provides a frequently needed guide for the direction of subsequent effort. It informs the investigator whether an ambiguously interpretable raw correlation can be ascribed to the poor but improvable quality of his measures or whether definite conceptual differences between the measures being related are likely to be involved. In our search for tenable, nonredundant theories, I simply do not see how a virtue can be made of blurring separate sources of variance.

In their final remarks, Stephens and Crowne are concerned with the practical problems surrounding usage of the correction for attenuation. I acknowledge such problems may exist, and carefully noted this point in my 1963 paper where I suggested the informed evaluator could at least establish reasonable and instructive bounds on the adjustment. The sophisticated "empirical integrator" will not hang his full set of conclusions on the precision of an estimate of a particular reliability coefficient. He will, however, recognize the operation of the attenuation principle and will make the order-of-magnitude adjustments required. In so doing, he will achieve a more accurate perspective on the state of the field and the payoff directions of subsequent effort than would result from neglect of this influence.

REFERENCES

Block, J. The equivalence of measures and the correction for attenuation. *Psychol. Bull.*, 1963, 60, 152–156.

Crowne, D. P., & Stephens, M. W. Self-acceptance and self-evaluative behavior: A critique of methodology. *Psychol. Bull.*, 1961, 58, 104–121.

Stephens, M. W., & Crowne, D. P. Correction for attenuation and the equivalence of tests. *Psychol. Bull.*, 1964, 62, 210–213.

(Received February 27, 1964)

VOL. 62, NO. 4 OCTOBER 1964

Psychological Bulletin

SIZE CUE TO VISUALLY PERCEIVED DISTANCE

WALTER C. GOGEL

Civil Aeromedical Research Institute, Oklahoma City

Evidence indicates that both perceived size S′ and retinal size θ are involved in both the relative and familiar size cue to relative depth. There is also evidence to indicate that the familiar size cue cannot be subsumed under the relative size cue. But both types of size cues can be included under the concept of perceived size per unit of retinal size (S'/θ). It is asserted that the perceived depth between objects, as determined by the size cue, is a function of the value of S'/θ associated with each of the objects. Evidence relevant to specifying the relation between values of S'/θ and perceived depth is evaluated with respect to methodological problems involved in the use of comparison fields.

The size cue is a frequent and important determiner of apparent distance. Two types of size cues will be considered. These are the relative size cue and the familiar size cue. Some recent evidence relevant to the operation of the relative and familiar size cue to distance will be examined and evaluated with respect to methodological problems involved in the measurement of apparent distance. A variable under which both the relative and familiar size cue can be subsumed will be identified. Experimental evidence concerning the form of the relation between this variable and perceived distance will be considered.

SIZE CUE TO RELATIVE DISTANCE

Relative Size Cue to Relative Distance

If two objects which are identical except for the size of their retinal image are presented in a situation in which all other distance cues are removed, the object with the smaller visual angle will appear to be the more distant object. This is the case of a relative retinal size determining a perceived depth between objects. It is as though the observer (O) assumes that the two objects are physically the same size and therefore perceives the retinally smaller object as being farther away.

Some of the terms used in this example and throughout this paper are labeled in Figure 1. Figure 1 is a perspective schematic drawing of three rectangles (e, f, and g) located at distances D_e, D_f, and D_g from O. The physical widths of the rectangles are labeled S_e, S_f, and S_g, with θ_e, θ_f, and θ_g the respective visual angles. Since retinal size and visual angle are usually proportional, θ_e, θ_f, and θ_g are also measures of the retinal size of the widths of the rectangles. The visually perceived distances of Rectangles e, f, and g from O (to be called perceived absolute distances) will be specified as D'_e, D'_f, and D'_g, respectively. The physical depths between the several rectangles are d_{ef}, d_{fg}, and d_{eg}, while the visually perceived depths (perceived relative distances) will be called d'_{ef}, d'_{fg}, and d'_{eg}, respectively. The visually perceived sizes associated with S_e, S_f, and S_g will be called S'_e, S'_f, and S'_g,

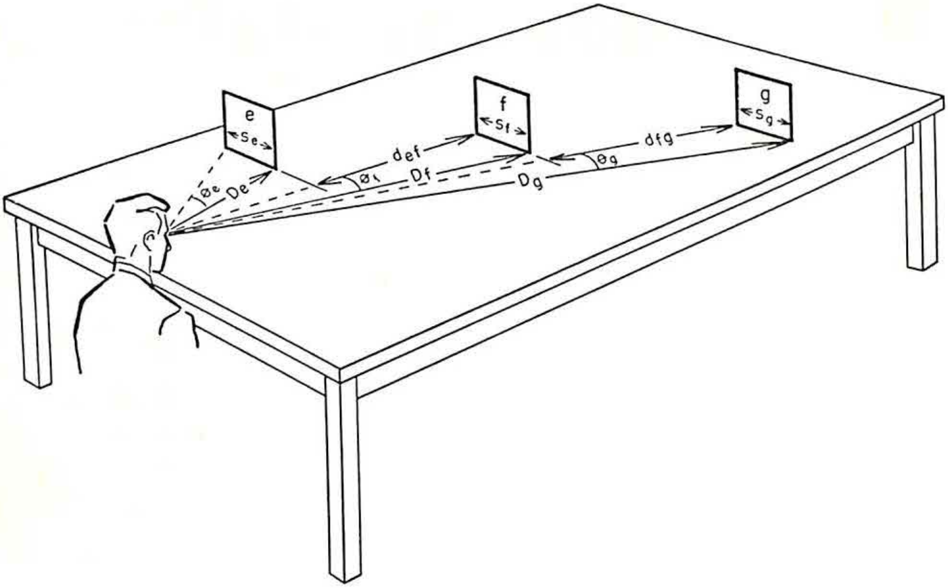


FIG. 1. A schematic perspective drawing useful in considering the operation of size as a cue to distance. (This figure has appeared previously as Figure 1 in Gogel, 1964b.)

respectively. In the particular example considered in the previous paragraph, Rectangles *e* and *f* are the only objects present and are viewed monocularly with the remainder of the field of view totally dark such that all distance cues are absent except that associated with the difference in the retinal sizes of the rectangles. Under these conditions, neither rectangle by itself would be perceived as having a unique size. However, the rectangles are assumed by *O* to be of the same size, and therefore, $S'_e = S'_f$. If D_f is greater than, equal to, or less than D_e, and $S_e = S_f$, it follows that θ_f is less than, equal to, or greater than θ_e, respectively. When $S'_e = S'_f$, and $\theta_e > \theta_f$, Rectangle *f* is perceived as more distant than Rectangle *e* (d'_{ef} is positive). When $S'_e = S'_f$ and $\theta_e < \theta_f$, Rectangle *f* is perceived as less distant than Rectangle *e* (for consistency, d'_{ef} should be considered as negative) and when $S'_e = S'_f$ and $\theta_e = \theta_f$, the two rectangles are perceived to be at the same distance (d'_{ef} is zero).

There is evidence that the simultaneous visual presentation of two (or more) similar but nonfamiliar objects such as Rectangles *e* and *f* in Figure 1, in the absence of other depth cues, will result in a perceived depth d'_{ef}, with the retinally smaller of the two objects appearing to be the more distant object. The term nonfamiliar is used here to indicate that there are no familiar characteristics associated with the objects which would specify a perceived size. For example, although rectangles are shapes which are frequently encountered, there is no unique size which would normally be associated with them. It is reported (Kilpatrick, 1961) that a simultaneously presented series of lines of different retinal sizes whose centers are all at the same height will appear at different distances with the smaller retinal sizes appearing more distant. The same phenomenon has been quantitatively demonstrated, for example, for a series of trapezoids (Gogel, 1954), for a pair of

squares (Epstein & Baratz, 1964, Experiment I), and for a pair of discs (Carr, 1935, pp. 262–263; Updegraff, 1930), etc. (cf. Ittelson, 1960, p. 70). Similarly, two balloons of different visual angle but of the same brightness will appear at different distances with the perceived depth between them changing with continuous variation in the size of one (Kilpatrick, 1961).

The relative size cue to relative depth will be effective when the objects (*e*, *f*, and *g* of Figure 1) are presented successively, as well as simultaneously, to the same *O* (Carr, 1935, pp. 260–261; Gogel, Hartman, & Harker, 1957; Over, 1963). The variability in the resulting perceived depth (d'), however, might be expected to increase as the time interval between the presentation of the two objects increased. The perception of a change in depth as a function of a continuous change in visual angle has also been demonstrated for a single nonfamiliar or familiar object (Ittelson, 1951b; Smith, 1951, 1952b, 1955; cf. Ittelson, 1960, p. 71). The evidence seems clear. Objects of the same shape with visual angles which differ either successively (as in the case of the same object of different retinal sizes presented at different times to the same *O*) or simultaneously (as in the case of presenting two similar objects of different retinal size at the same time) or continuously (as in the case of the object whose visual angle changes while *O* is observing it) will result in a perception of relative depth.

Relative values of S' as well as θ are necessary in order for the relative size cue to result in a perceived distance. There is evidence (Hochberg & McAlister, 1955) that two simultaneously presented rectangles or two simultaneously presented circles of different retinal sizes can produce an apparent depth but, that the simultaneous presentation of a rectangle and circle of different retinal sizes does not. Clearly, some perception (S'_e and S'_f) of the relative sizes of the objects (*e* and *f*) is necessary. Neither S'_e nor S'_f must necessarily be perceived as being a particular size, but some perception of the relation between the sizes is necessary. Also, while a difference in shape between two objects tends to destroy the perception of relative depth, it is not necessary that the two objects be identical in all characteristics except that of retinal size. It has been found, for example, that a depth perception can occur as the result of the simultaneous presentation of two squares of identical shape but different color (Epstein & Baratz, 1964, Experiment I). It is also unnecessary to assume that S'_e must equal S'_f in order for the relative size cue to result in a perception of relative depth. This can be shown by the results from a portion of a study by Over (1963) in which the perceived sizes and perceived distances of successively presented, diamond-shaped squares of various retinal sizes were measured under visually reduced conditions. The present author concludes that the results from this study by Over indicate that perceived distance can vary between presentations as a function of the changes in retinal size (see the open circle data of Figure 2b, p. 228) even though the perceived sizes of the squares were not always equal in the different presentations.

In summary, the experimental evidence is in agreement with the following description of the essential processes involved in the relative size cue to relative distance: (*a*) The relative size which is referred to is a relative retinal size[1] and

[1] The expression "relative size judgments" (Gogel et al., 1957) or "relative size" (Baird, 1963) has sometimes been used to refer to the perception of the relative sizes of objects presented either simultaneously or successively, regardless of whether the perceived size occurred between similarly shaped objects or not. Instead, the expression "relative size cue" in this paper follows the meaning used by Hochberg and Mc-

is a difference (or equality) in retinal sizes occurring from the simultaneous or successive presentation of similar objects. If the retinal sizes are different, a perception of depth between the objects will occur with the object having the retinally smaller size appearing more distant. If the retinal sizes are the same, the objects will appear equidistant. (*b*) Relative retinal size, by itself, is not sufficient to produce a perception of relative depth. Some similarity of shape between the objects must be present, and (*c*) must result in some perception of the relative sizes of the objects, with the usual perception being that of size equality. It is clear that relative values of S' as well as relative values of θ are involved in the relative size cue to relative depth.

Familiar Size Cue to Relative Distance

Familiar (or assumed) size as a cue to relative distance can also be considered with the aid of Figure 1. Suppose that Objects *e* and *f* are presented simultaneously and are monocularly viewed with the remainder of the field of view totally dark. Suppose also that Objects *e* and *f* are familiar objects, for example, two playing cards. Since $\theta_e > \theta_f$, Object *e* will be perceived as less distant than Object *f* (Gogel, 1956a; Gogel & Harker, 1955; Ittelson, 1951c). In this example, however, it is not necessary to resort to familiar size to explain the resulting perceived depth. An explanation in terms of relative size would be equally plausable (Hochberg & Hochberg, 1952, 1953; Hochberg & McAlister, 1955). Suppose, however, that Object *f* appears to be a similarly shaped but physically smaller object than a playing card, for example, a matchbox, with Object *e* continuing to be seen as a playing card. In this latter example, Object *f* might appear to be less distant than Object *e* even though $\theta_e > \theta_f$. According to the usual explanation of the familiar size cue, each of the objects (the matchbox and the playing card) should appear at a distance position which a normal-sized object of that particular category would have to occupy to produce that particular retinal size (Ittelson, 1951c). In the latter example since the objects are of similar shape and $\theta_e > \theta_f$, Object *f* should appear behind Object *e* according to the relative size cue but in front of Object *e* (for some values of $\theta_e > \theta_f$) according to the familiar size cue. Thus, it has been argued (Epstein & Baratz, 1964; Ittelson, 1960, p. 71) that the relative size cue and the familiar size cue in such situations are in opposition to each other.

The question as to whether the cue of familiar size has been demonstrated to be distinct from the cue of relative size has been answered in the affirmative by Ittelson (1951c) and in the negative by Hochberg and Hochberg (1952, 1953) and by Hochberg and McAlister (1955). A recent experiment by Epstein and Baratz (1964, Experiment II) is pertinent to this issue. In the experiment by Epstein and Baratz, perceived depth was measured between objects representing a dime, a quarter, and a half dollar presented in pairs. Each coin subtended three possible retinal sizes such that the relative and familiar size cues could be placed in agreement or in opposition. The results of the study indicate that the perceived depth between the pairs of coins when the two cue systems were in opposition was always in agreement with the familiar, not the relative size cue. It seems reasonable to conclude that the familiar size cue to relative depth cannot be subsumed under the relative size cue to relative depth.

Alister (1955), that is, the relative size cue as used in this paper is the relative depth cue which occurs from observing two (or more) similarly or identically shaped objects of different retinal size.

Specification of the Significant Variable in the Size Cue to Relative Distance

From the previous discussion, it follows that in the perception of relative depth which results from the cues of relative and familiar size, the important variables are the relative values of S' and θ. In both the relative and familiar size cue, θ is determined by the physical size and physical distance of the stimulus objects. In the relative size cue, relative values of S' are determined by the perception that the identically shaped objects are identical or similar in size. In the familiar size cue, values of S' are determined by the size associated with each of the familiar objects. A difference in apparent size by itself does not determine a perceived relative distance. For example, in the case of the relative size cue, two identically shaped objects of different retinal sizes appear to be at different distances because they appear to be the same size. Also, perceived depth from the size cue is not determined solely by the relative retinal sizes of the objects. For example, as shown by Epstein and Baratz (1964), two objects of the same retinal size will appear to be at different distances because they appear to be of different sizes. It follows from such considerations that S' and θ considered jointly determine the perceived relative depth resulting from the size cue for both relative and familiar size. The point of view of this paper is that the significant variable in the size cue to relative depth is the relative value of S'/θ for each of the objects being considered.

The term S'/θ is the perceived size per unit of retinal size of the particular frontoparallel object being considered. When $S'_f/\theta_f > S'_e/\theta_e$, Object f is perceived to be more distant than Object e. When $S'_f/\theta_f = S'_e/\theta_e$, Objects e and f are perceived to be equidistant and when $S'_f/\theta_f < S'_e/\theta_e$, Object f is perceived to be less distant than Object e. It is asserted that the perceived depth d'_{ef} is a function of the values of S'/θ for the two objects. To apply the concept of S'/θ to the relative size cue, consider a situation in which the similarity of shape of the two nonfamiliar objects results, for example, in the perception that $S'_e = S'_f$. Thus, that object in the pair of objects which has the smallest value of θ will have the largest value of S'/θ and will appear to be the more distant object. To apply the concept of S'/θ to the familiar size cue consider a situation in which, for example, two objects (e and f) are of the same shape and retinal size ($\theta_e = \theta_f$) and are perceived to be a matchbox and a playing card, respectively. Since experience indicates that a playing card is larger than a matchbox, $S'_f > S'_e$ and thus $S'_f/\theta_f > S'_e/\theta_e$. Therefore, the playing card (f) will be perceived to be more distant than the matchbox (e). Familiar size and similarity of shape are merely different methods of determining values of perceived size (S'). Other factors such as assumptions (Hastorf, 1950) and sometimes, but not always, special experience (Epstein, 1961; Smith, 1952a) are also determiners of perceived size. Therefore, the familiar size cue and the relative size cue are never in opposition. It is the perceived sizes of the objects regardless of the manner in which these perceived sizes are produced which together with the retinal sizes results in a perceived depth. The concept that the relative values of S'/θ for the several objects determine the resulting perceived depth between the objects subsumes both the relative and familiar size cues into one system.

When Object e is of different shape than Object f, the concept of the relative size cue does not apply. If the objects are also familiar objects, however, the concept of the familiar size cue does apply. It has been asserted that relative depth

judgments using the familiar size cue can be made under such conditions (Ittelson, 1951c, 1953). This assertion would also be made from the viewpoint taken in the present paper that variations in S'/θ between objects is the significant variable in the perception of relative depth from size cues. Two objects which differ in S'/θ, in the absence of other distance indicators, will be perceived at different distances. But, the concept of S'/θ is not only applicable to situations in which the familiar objects are of the same or similar shape. The concept of S'/θ as the significant variable in the size cue to relative depth will apply to irregularly shaped objects such as trees and animals as well as to geometric figures such as playing cards or circles. The application of S'/θ to differently or irregularly shaped objects also can be illustrated by Figure 1. Suppose in Figure 1 that Objects *e*, *f*, and *g* are increased in size so that they represent frontoparallel planes rather than objects of limited size. Furthermore, suppose that a house is pictured as located in Plane *g*, a horse and tree in Plane *f*, and a man in Plane *e*. Also suppose that these three planes are apparent frontoparallel planes rather than physical frontoparallel planes, and that no distance cue other than that of the size cue is available to determine the apparent depth between the three planes and the various objects in different planes. According to the concept that S'/θ is the significant variable in the size cue, the objects in the different apparent frontoparallel planes are in different apparent frontoparallel planes because they have different values of S'/θ. Also, the objects in the same apparent frontoparallel plane (and all portions of an object in the same apparent frontoparallel plane) are in the same apparent frontoparallel plane because they have the same value of S'/θ. For example, suppose that two marks made on the trunk of the tree in Plane *f* are separated by 1 degree of visual angle and are perceived to be separated by 1 foot. In this case, the S'/θ value is 1 foot per degree. A separation of .5 degree of visual angle anywhere else on the tree or on the horse, or anywhere else in Plane *f* should appear to be 6 inches in length, giving thereby the same value of S'/θ. It follows that a constant value of S'/θ applies to an entire apparent frontoparallel plane including all the objects and interspaces within this plane regardless of their irregularity. Apparent frontoparallel planes having greater values of S'/θ will appear more distant, while those with lesser values of S'/θ will appear less distant. It is asserted that the different values of S'/θ between the objects in different apparent frontoparallel planes or between different planes will determine the perceived depth between these objects or planes. In a situation in which the size cue to relative depth is the only distance cue present, it is hypothesized that all the perceived depths within this visual field are immediately specified by specifying the S'/θ values of the different objects.

Size Cue to Absolute Distance

In the discussion involving Figure 1, the perceived distance D' of an object from O was called a perceived absolute distance to distinguish it from d' the perceived depth between objects (perceived relative distance). The size cue discussed previously is the size cue to perceived relative distance, not the size cue to perceived absolute distance. It is usually assumed (at least implicitly) that the reason why familiar size is a cue to perceived relative distance is because it is a cue to perceived absolute distance. It is supposed that experience will result in perceived absolute distances being associated with retinal sizes of a familiar object such that for any particular retinal size an appropriate absolute distance will be perceived (Ittelson, 1951c). From this point of view, a perceived relative

distance is merely the difference between two perceptions of absolute distance. If this point of view is correct, the psychophysics of familiar size as a cue to relative distance is specified when the psychophysics of familiar size as a cue to absolute distance is known. Evidence relevant to this point of view will be examined in this portion of the paper.

A number of studies involving the so-called "thereness-thatness" type of apparatus have been interpreted as supporting the notion that the retinal size of a familiar object can determine a perceived absolute distance (Epstein, 1963; Hastorf, 1950; Ittelson, 1951c; Ittelson & Ames, 1950). In these studies the objects with the familiar (or assumed) size (the experimental objects) were presented in a visually reduced (experimental) field. The perceived distance of the experimental object was measured by comparing the apparent position of the experimental object with respect to that of a comparison object presented in a comparison field. The comparison object was either part of a visually well-structured field or was itself a familiar object. In either case, visual size comparisons could occur between the experimental object and objects in the comparison field. It has been asserted (Gogel et al., 1957) that the measures obtained under these conditions have not been demonstrated to be measures of perceived absolute distance. The same assertion has been made recently (Adelson, 1963) in a study in which interactions between the experimental and comparison fields have been reported.

The kind of judgment which may occur when the "thereness-thatness" type of apparatus is used can be analyzed with the aid of Figure 1. Suppose that a variety of cue systems are available in the field of view represented by Figure 1, such that any object physically located in this field of view in any physical frontoparallel plane (e.g., Objects e, f, or g) is correctly perceived with respect to both its size and distance. Also suppose that by means of appropriately placed mirrors, a playing card is made to appear in the visual field of Figure 1, with no cues present to localize the card in distance except its familiar size. The field of view of Figure 1 corresponds to the comparison field in the thereness-thatness apparatus while the playing card corresponds to the experimental object. The playing card will have a perceived size S' as determined by its familiar characteristics and also will have a retinal size θ. Thus, the playing card will have a particular value of S'/θ as will every object in the perceptually well-structured comparison field. If, as this paper asserts, S'/θ is the significant variable in the size cue to relative depth, the playing card will appear in that apparent frontoparallel plane in the comparison field which has the same value of S'/θ as does the playing card. All that is required of O for this perception to occur is a null comparison (a judgment of equality) of the S'/θ values of objects in the comparison field with respect to the S'/θ value of the playing card. Entirely incidental to this process is the question of how far the playing card appeared to be from O. No judgment of perceived absolute distance is required. The physical distance from O to the position of the plane (or to the position of objects in the plane) at which the card appears can be measured as is done with the thereness-thatness apparatus. But, if the point of view expressed in this paper is correct, this measure has nothing to do with O's task and does not represent a perception of the absolute distance of the card. Similar comments apply when the method of measuring apparent distance involves adjusting the comparison object or the experimental object to apparent equidistance with each other. Furthermore, since visual comparisons between S'/θ values can occur between presentations

separated in time, this objection to the use of comparison fields is not removed by presenting the experimental and comparison fields successively rather than simultaneously. It is suggested that the results from the thereness-thatness apparatus demonstrate that experimental and comparison objects with the same S'/θ values appear equidistant, but it is further suggested that the use of the thereness-thatness apparatus does not result in measures of perceived absolute distance. A similar type of objection has been stated previously (Gogel et al., 1957) by noting that the measurement of the perceived absolute distance of a familiar object requires that a test situation be used in which only the familiar object is presented visually to *O*. A second requirement for the measurement of perceived absolute distance is not specific to the thereness-thatness type of apparatus. It is asserted that if different retinal sizes of the familiar object are used these must be presented to different *O*s. Comparisons can occur between successive presentations of different retinal sizes of the same or different familiar objects with the result that observer judgments which are actually judgments of relative distance may be interpreted by the experimenter (*E*) as being judgments of absolute distance.

From the point of view of the above criteria, there have been few adequate tests of whether familiar size is a cue to absolute distance. The study by Gogel et al. (1957) was designed to meet the above requirements. Kinesthetic judgments and verbal reports were used to measure the perceived absolute distance to singly presented playing cards in visually reduced situations using different groups of *O*s for the first presentations of different retinal sizes. Only the responses to the first presentations were considered to represent perceived absolute distances. Response changes as a function of successive presentations were considered to measure perceived relative distance. The results from this study indicate that the retinal size of a familiar object in isolation is not an effective cue to perceived absolute distance and also that the perception of relative depth which occurred between successive presentations of the cards occurred independently of any perception of absolute distance. A recent study by Epstein (1963), using figures of coins of three different denominations but constant retinal size as familiar objects and with a different group of *O*s viewing each coin, seems to support the point of view that familiar size is a cue to perceived absolute distance. However, since the judgments were made using a thereness-thatness type of apparatus, the experimental results do not necessarily measure perceived absolute distance but, instead, probably indicate the ability of *O*s to perceive that experimental and comparison objects which have the same value of S'/θ are equidistant.

A recent study by Baird (1963, Experiment II) meets the criteria discussed above. Three groups, with 10 *O*s in each group, made verbal estimates of the distance of a luminous rectangular strip, with each group presented with a different retinal size. The *O*s were informed that the figure was the same size as a 12-inch ruler. The figure which was in a darkened room was always physically at the same distance (25 feet) and was observed monocularly. The different groups gave average results which were close to the physical distances at which a 12-inch ruler would have to be placed to subtend the particular visual angles. This study supports the point of view that the retinal size of an object of a particular assumed size can be a determiner of perceived absolute distance. It was noted, however, that the intersubject variabilities for the distance estimates were large (standard deviations of 5.6, 7.3, and 42.6 feet associated with the

mean distance estimates of 14.3, 24.0, and 48.2 feet, respectively). It is unlikely that the perception of absolute distance from assumed or familiar size occurs with sufficient precision to provide an explanation for the perception of relative depth from assumed or familiar size.

It was pointed out in the study by Gogel et al. (1957) and also in the study by Baird (1963) that, since size comparisons can occur over time, it is difficult to measure perceived absolute distance uncontaminated by successive size comparisons. Probably, the time separation between the previous visual experience and the judgments in the test situation should be as long as possible. It is, however, important to determine the limiting conditions under which judgments of absolute distance can occur to further examine if indeed perceived absolute distance can provide a basis for the explanation of perceived relative distance (Gogel, 1963). The study by Gogel et al. (1957) directly indicates that the perceived relative distance resulting from size comparisons can occur independently of perceived absolute distance. If this conclusion is correct, it follows that the psychophysics of the perception of relative distance resulting from the size cue needs to be reexamined.

Measurement of Perceived Relative Distance with a Comparison Field

It is probable that the thereness-thatness apparatus does not measure the perceived absolute distance of familiar objects. It is equally unlikely that it measures the perceived relative distance of familiar objects. Consider the case in which two familiar objects are viewed monocularly in a visually reduced experimental field with the perceived depth between the two objects measured by the physical distance between their apparent positions in a full-cue comparison field. Each of the familiar objects in the experimental field would consequently appear at the distance of that portion of the comparison field which has the same value of S'/θ as the familiar object. Again, only apparent equidistance judgments would be required by *O*. It follows that the physical distance between the positions of the objects in the comparison field which appeared equidistant to the familiar objects in the experimental field would not be a valid measure of the perceived distance between the experimental objects. According to this point of view, *O* has matched S'/θ values between each experimental object and portions of the comparison field. The perceived depth between the two experimental objects has not been represented in *O*'s performance and consequently has not been measured.

Attempts to measure the perceived relative depth between two nonfamiliar objects by means of a comparison visual field also introduces problems. Suppose, in the previous example, that two luminous squares of different retinal size are used instead of the familiar objects. Under these conditions, the retinally smaller square will appear to be the more distant square. But, what will be the perceived depth between the squares? The squares by themselves provide no scale or metric by means of which such judgments can be made. An *O*, if asked to make a relative depth judgment without the comparison field present, may do so. But it is doubtful whether such a judgment can be meaningful. However, Epstein and Baratz (1964) have found that *O*s can adjust two objects in a comparison field to the apparent radial distance of each of the two simultaneously presented, experimental squares of different retinal size. The distance between the two adjustments was interpreted to be a measure of the perceived depth between the squares. Evidently, the comparison field introduced a metric which was not in the original perception. Perhaps even

more difficult to understand is the finding that nonfamiliar experimental objects such as ink blots, diamond-shaped figures (Ittelson, 1951c), or discs (Epstein, 1961, Experiment II) presented singly will have an apparent localization with respect to the comparison field which changes systematically as a function of their retinal size. In this latter case, however, it has been found (Epstein, 1961, Experiment III) that the systematic change is considerably reduced when different groups of *O*s are used with the different retinal sizes. When the same *O*s are used with different retinal sizes, it seems that a perception of relative not absolute distance is being measured as a consequence of the successive presentations. A possible explanation of how a perception of relative distance can result from a retinal size difference between successively or simultaneously presented nonfamiliar objects as a consequence of the presence of a comparison field is as follows: It is very difficult to eliminate all extraneous factors (factors other than the cue system being studied) which can relate the depth position of one or more of the experimental objects with respect to the comparison field. It has been found, for example, that the accommodative differences between the experimental object and the comparison field sometimes can determine the apparent depth position of the experimental object (Gogel, 1961). Also, even when the difficult task of completely eliminating extraneous cues is accomplished, factors such as the equidistance tendency (Gogel, 1956b) probably would result in some localization of the experimental objects with respect to the comparison field. Assume that one of the discs, or squares, or ink blots, as a consequence of extraneous factors, will appear to be located somewhere in depth with respect to the comparison field. It follows that the S'/θ value associated with this apparently equidistant portion of the comparison field will become the S'/θ value of the object. In other words, the disc, or square, or ink blot will acquire a perceived size. Then, as a result of shape identity, the other disc, or square, etc., (either simultaneously or successively presented) will acquire the same perceived size. Each of the pair of squares, etc., will thus acquire an S'/θ value and, therefore, a specific perceived relative depth. Consequently, the perceived relative depth measured between nonfamiliar objects using a comparison field is probably not independent of the comparison field. If the above explanation is correct, however, the perceived depth which is measured, in this case, probably is less contaminated by the measuring (comparison) field than is likely in the case of familiar objects.

The usual purpose of using comparison fields in the investigation of size as a cue to distance is to measure the perceived absolute distances of experimental objects or the perceived depth between them *as these distances would have appeared had no comparison field been present.* From the above discussion, it is likely that this purpose is usually not achieved.

Possible Relations between Perceived Distance and S'/θ

Application of the Size-Distance Invariance Hypothesis

A conclusion of this paper is that the perceived relative depth between objects resulting from size cues is some function of the S'/θ values of the objects. It is asserted that the variable S'/θ applies to both relative and familiar size cues so that these seemingly different types of size cues involve the same basic process. The task remains, however, of determining the specific relation between d' and values of S'/θ. One suggestion for this relation comes from the size-distance invariance hypothesis (cf. Kilpatrick &

Ittelson, 1953) of which perceived size as a cue to distance is a particular case (Epstein, Park, & Casey, 1961). The size-distance invariance hypothesis can be expressed (Gogel, 1964b; Kilpatrick & Ittelson, 1953) as

$$D' = \frac{1}{K_1}\frac{S'}{\theta}, \qquad [1]$$

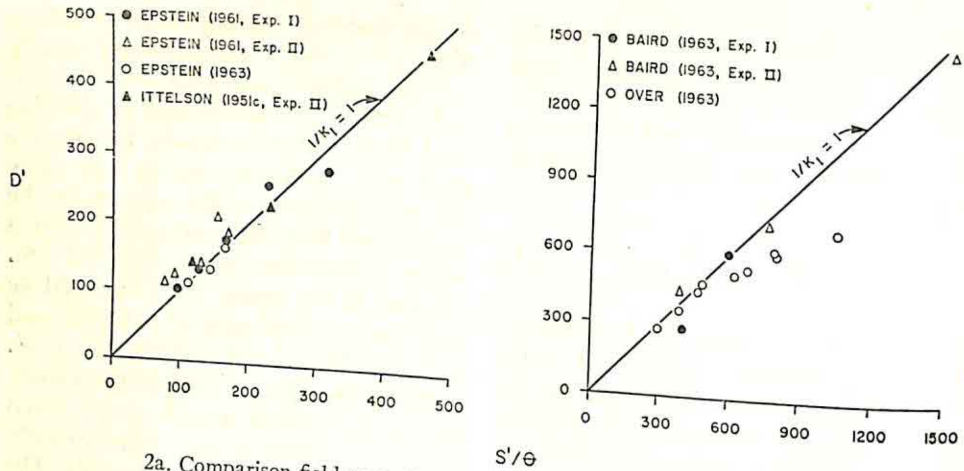
where K_1 is a constant for a particular observer or situation. Equation 1 states that a particular value of retinal size (θ) and a particular value of perceived size (S') determines a particular perceived absolute distance (D'). Also, according to Equation 1,

$$d'_{ef} = \frac{1}{K_1}\left(\frac{S'_f}{\theta_f} - \frac{S'_e}{\theta_e}\right), \qquad [2]$$

where $d'_{ef} = D'_f - D'_e$. From the previous discussion, the use of a comparison field, for the measurement of the perceived distance of familiar experimental objects results in apparent equidistance adjustments as determined by equal values of S'/θ of the experimental and comparison objects. A review of the available evidence will demonstrate that when a comparison field is used the resulting data usually are in agreement with Equation 1. However, it will also be asserted that such agreement would occur whether Equation 1 is valid or not.

Suppose that a double-sized playing card (17.8 cm. high) is presented at 304 cm. in an experimental field and its apparent distance measured with a comparison field. The double-sized card usually will appear to be 8.9 cm. high (its normal height) and will subtend a visual angle of .058 radians. It follows that the S'/θ value of the playing card is 152 cm/radian. The comparison field is a full cue situation such that the objects comprising the comparison field will be perceived to be their physical size. Therefore, an object, for example, a square, 8.9 cm. high, at 152 cm. in the comparison field will appear to be 8.9 cm. high, will subtend a visual angle of .058 radians, and will have an S'/θ of 152 cm/radian. Since the square and the playing card have equal S'/θ values, the playing card will appear to be at the same distance position as the square in the comparison field. As is the practice when using comparison fields, the 152 cm. distance of the square from O would be taken to be a measure of the perceived absolute distance of the playing card. Actually the playing card in the experimental field might have been perceived by O at any or no absolute distance without changing the measured result. The unwarranted conclusion would be made that the double-sized playing card physically located at 304 cm. actually appeared to O to be 152 cm. from himself. From the results of such measurements, the unwarranted conclusion would be drawn that familiar objects of particular retinal sizes will be perceived to be at distances which normal-sized objects of that familiar category would have to occupy in order to produce the particular retinal sizes. This is also the conclusion which would be predicted from Equations 1 or 2 when $1/K_1 = 1$. Therefore, the use of comparison fields should result in Equations 1 or 2 being satisfied regardless of the validity of these equations.

In Equation 1, measurements of S', D', and θ should reflect the perceived size, the perceived distance, and the visual angle, respectively, of the experimental object. However, according to the point of view of this paper, when the "thereness-thatness" apparatus is used, the experimental object will appear equidistant with that object in the comparison field which has the same value of S'/θ as does the experimental object. Furthermore, the rationale of using comparison fields requires that the physical distance of this apparently equidistant object in the comparison field be taken as the perceived



2a. Comparison field present.

2b. Comparison field absent.

FIG. 2. The relation between perceived distance (D') and perceived size per unit of retinal size (S'/θ). (The stimulus objects used in the different experiments were either familiar objects [Epstein, 1961, Experiment I; Epstein, 1963; Ittelson, 1951c, Experiment II], objects with induced apparent sizes [Baird, 1963, Experiments I and II], or nonfamiliar objects [Over, 1963]. The terms D' and S' are in centimeters and θ [visual angle] is in radians.)

distance of the experimental object. Therefore, S' and θ refer to events in both the comparison and experimental fields while D' is measured by D in the comparison field. Equation 1 can be put in the form:

$$\frac{S}{D} = \frac{1}{K_1}\frac{S'}{D'}, \qquad [3]$$

where $S/D = \theta$ in radians. But if D' is assumed to be measured by D in the comparison field, it follows from Equation 3 that the only requirement in order for $1/K_1$ to equal unity in Equations 1 and 2, is that S' is proportional to S in the comparison field. This condition might be termed the condition of veridicality of perceived size in the comparison field, that is, perfect size constancy in the comparison field. It follows that, if perfect size constancy occurs in the comparison field, and if the judgmental process involves only the comparison of S'/θ values between experimental and comparison objects, the data obtained from using the comparison field should have the slope $1/K_1 = 1$, when D' is plotted against S'/θ with θ expressed in radians. If this reasoning is correct, the results from using a comparison field are likely to measure the amount of size constancy in the comparison field rather than the perceived absolute distance of the experimental object. It will be noted that a value of $1/K_1 = 1$ in Equations 1 and 2 would also be expected if the familiar object would be perceived at a distance which a normal-sized object of that particular category would have to occupy to subtend the particular retinal size, that is, if the retinal size of a familiar object were a veridical cue to perceived absolute distance.

Figure 2a shows the relation between D' and S'/θ from three experiments in which a thereness-thatness type of apparatus was used. The solid line in Figure 2a represents the results which would be expected if the data fitted Equation 1 exactly and if $1/K_1$ were unity. However, any series of data points which describe a linear function passing through the origin of Figure 2a would satisfy Equation 1 with the slope of the

function determining $1/K_1$. The data points marked by filled triangles in Figure 2a represent the average results from successively presenting three different retinal sizes of a playing card to the same group of observers in a study by Ittelson (1951c, Experiment II). In order to calculate S'/θ for this study, it was assumed that S' was 8.9 cm. (the height of a normal-sized playing card). The average data are very close to the line representing $1/K_1 = 1$. The data points marked by filled circles in Figure 2a give the average results from a control group of *O*s in an experiment by Epstein (1961, Experiment I) in which playing cards of different angular sizes were presented successively with measurements taken of both apparent size and apparent distance. In spite of the factor of familiar size, the cards were not always perceived to be of the same size. Nevertheless, the data reasonably fit the line representing $1/K_1 = 1$. From the viewpoint of the present paper, the results from both of these studies are interpreted to mean that the cards were perceived to be located at those distance positions in the comparison field at which the comparison and experimental objects had equal values of S'/θ and that, as might be expected, perceptions of size in the comparison field were essentially veridical. When the perceived size S' of the playing card in the experimental field differed from that expected from familiar size, the measured perceived distance was in agreement with the S'/θ value, not in agreement with familiar size. The data points marked by open circles in Figure 2a are the average results from the study by Epstein (1963) in which figures of different kinds of coins with different familiar sizes, but with constant retinal sizes, were presented singly to different groups of *O*s. Pinhole viewing was used to eliminate accommodation as a distance cue. Both the sizes and the distances of the coins were judged. The judged sizes of the coins were close to those of normal-sized coins of the particular denominations. The average data lie near the solid line representing $1/K_1 = 1$ in Figure 2a.

Equation 1, with a value of $1/K_1 = 1$, also can be approximately satisfied when nonfamiliar objects are used in the thereness-thatness type of apparatus. This is indicated by the results obtained with nonfamiliar objects in the study by Ittelson (1951c) and also by the open triangles marking data points in Figure 2a. The open triangles marking data points in Figure 2a are the average results from an experiment in which different discs of different colors and different angular sizes were presented one at a time to the same group of *O*s with observer judgments made of both the sizes and distances of the discs (Epstein, 1961, Experiment II).

It seems that the data of Figure 2a are in agreement with Equation 1, since essentially the data points seem to lie on a straight line passing through the origin. But, the data do more than satisfy Equation 1, in general. Specifically, the data lie close to the line specified by $1/K_1 = 1$. These results would be expected if, as this paper asserts, the *O*s (regardless of their intentions) were matching S'/θ values between the experimental objects and objects in a veridically perceived comparison field instead of judging the absolute distances of the experimental object. Either this interpretation or Equation 1 is adequate for all the results shown in Figure 2a. The interpretation that the standard objects were perceived at absolute distances, consistent with normal-sized objects of the particular retinal sizes, is appropriate for all but the data points marked by open triangles in Figure 2a.

If the use of comparison fields is not an appropriate method of testing Equations 1 and 2, and if Equations 1 and 2 are not valid, methods of measuring per-

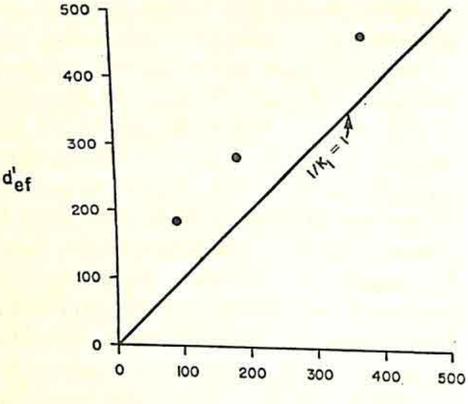
ceived depth from size cues which do not use comparison fields should give results in disagreement with these equations. The study by Gogel et al. (1957) in which playing cards of different retinal size were presented one at a time to four different groups of *O*s is pertinent to this problem. Perceived distance was measured by kinesthetic judgments and verbal reports rather than by means of a comparison field. The perceived absolute distance of the cards was indicated by the responses to the first presentations of the card. Perceived relative distance was indicated by the response changes between successive presentations of the cards to the same *O*s. In applying the results from this study to Equation 1, the perceived size S' of the playing cards can be assumed to be the height of a normal-sized card (8.9 cm.). The results from neither the first nor the successive presentations support Equation 1. In the former case, the perceived absolute distance of the cards did not differ significantly as a function of their retinal size (θ). In the latter case, a plot of the average values of D' against S'/θ, if extended, would not pass through zero as Equation 1 requires.

Results from other studies in which comparison fields were not used are shown in Figure 2b. In a study by Over (1963), four *O*s made 90 verbal judgments of the size and distance of diamond-shaped figures (squares) of different physical sizes presented successively, at various distances, under reduced viewing conditions, that is, with no size or distance cues present except the size cue between successive presentations. Following this, the same *O*s repeated the judgments under unrestricted viewing conditions, that is, conditions in which many size and distance cues were available. Since the same *O*s were used for the successive presentations, the judged distances in the reduced cue situation probably were not perceived absolute distances. However, the differences between pairs of distance judgments can be interpreted as representing perceived relative distances. No perceptual scale was available under the reduced viewing conditions and it is, therefore, unlikely that the magnitude of any single perceived relative distance was meaningful. However, changes in perceived relative distance as a function of changes in relative retinal size would be meaningful data. The results from the reduced viewing conditions are given by the open circles marking data points of Figure 2b. The solid line curve of Figure 2b again represents Equation 1 with $1/K_1$ equal to unity. The judgments made in the reduced viewing conditions are not in agreement with Equation 1. As shown by the data points marked by open circles, judgments made in the reduced viewing conditions appear to determine a curvilinear rather than a linear function. The judgments made under the unrestricted viewing conditions involved many cue systems in addition to that of size. The data from the unrestricted viewing conditions (not shown in Figure 2b) as expected, are close to the curve labeled $1/K_1 = 1$. The unrestricted viewing data are not pertinent when considering the function involved in perceived size as a cue to distance. However, judgments from the unrestricted viewing conditions do demonstrate that the method of judging size and distance used in the study was valid and reliable. It follows that the curvilinearity of the "reduced viewing" function cannot be attributed to a nonlinear scale inherent in the method of measurement and, therefore, the results from the reduced viewing conditions represent valid evidence against Equation 1 (and 2).

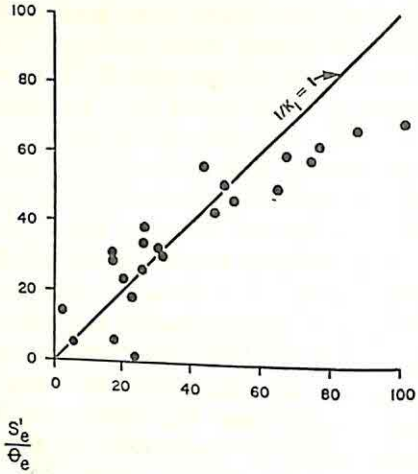
Average data from two experiments by Baird (1963) are shown by the data points marked by the filled circles and open triangles of Figure 2b. The data points marked by open triangles refer to the average verbal reports of the distance of a rectangular figure designated as the size of a 12-inch ruler (Experiment II). As discussed previously, different retinal

sizes of this figure were presented to different groups under reduced conditions of observation, with 10 *O*s in each group. It will be noted that the resulting curve is essentially in agreement with Equation 1, with $1/K_1 = 1$. The data points marked by filled circles in Figure 2b refer to the average results from Experiment I by Baird in which different assumed sizes were induced in a triangle of constant retinal size by first showing rectangles of differing retinal size with each rectangle designated as being 12 inches wide. Only the results obtained under objective instructions in Experiment I were used in Figure 2b. Size judgments were made by using a comparison triangle, and distance judgments were made verbally. Most likely, the data points marked by filled circles represent relative judgments of size and distance. Since the distance judgments for the middle and small retinal sizes were greater than the distance to the comparison object, it is likely that the comparison field could not completely determine the distance judgments. Clearly, the curve which would result from connecting the data points marked by filled circles does not fit Equation 1. Also, this curve, unlike the curve which would result from connecting the data points marked by triangles in Figure 2b, seems essentially nonlinear. The data points marked by the open triangles and filled circles involve perceived absolute and perceived relative distance, respectively. However, if the curve resulting from the perceived absolute distance data is linear, the curve resulting from the perceived relative distance data, according to Equation 1, should be linear also. The difference between the shape of these two curves from the study by Baird (Figure 2b) and the difference between the results from the first and successive presentations in the study by Gogel et al. (1957) can be interpreted in a similar manner. The results from both studies can be interpreted to mean that perceptions of relative distance from size cues are not always explainable in terms of differences between perceptions of absolute distance.

Only the data points marked by open triangles in Figure 2b support Equation 1. The data in Figure 2b which do not support Equation 1, also do not support Equation 2. However, there are several studies which directly have measured d' rather than D' as a function of size cues. An experiment which can be applied here is the experiment by Ittelson (1951b, Experiment II) in which three playing cards of different physical sizes were presented one at a time, and moved a constant distance toward and away from *O*, in a visually reduced experimental field. Using a comparison field, five *O*s judged the nearest and farthest positions of the movement for each of the physical sizes of the card. The physical distance between these judgments can be taken as d', θ can be calculated from the physical sizes and the physical distances of the near (e) and far (f) positions, and S' can be assumed to be the size of a normal playing card (8.9 cm. high). The results from plotting the average values of d' against differences in S'/θ are shown by the data points in Figure 3a. The straight line labeled $1/K_1 = 1$ again represents the results expected if $1/K_1$ in Equation 2 were unity. The curve which would result from connecting the data points of Figure 3a, although linear with a slope of approximately unity, would not pass through zero as Equation 2 would require. Either the data do not support Equation 2, or the discrepancy is a consequence of some interaction between the experimental and comparison fields, which effectively added a constant to the obtained results. In the study by Epstein and Baratz (1964, Experiment II), a comparison field was used to measure the perceived depth between pairs of simultaneously presented coins with each coin subtending one of three possible retinal sizes. Thirty-two *O*s adjusted two nonsense-form comparison figures to duplicate the apparent distance between each of the pairs of coins. The radial difference



3a. Apparent depth between terminal depth positions of moving playing cards (Ittelson, 1951b, Experiment II).



3b. Apparent depth between pairs of coins (Epstein & Baratz, 1964).

FIG. 3. The relation between perceived relative distance (d'_{ef}) and differences between perceived sizes per unit of retinal sizes ($S'_f/\theta_f - S'_e/\theta_e$). (The terms d'_{ef}, S'_e, and S'_f are in centimeters. The terms θ_e and θ_f are visual angles expressed in radians.)

between the adjustments in the comparison field for each pair of coins was considered to be the perceived depth d' between the pairs of coins. Figure 3b shows the obtained relation between d'_{ef} and $S'_f/\theta_f - S'_e/\theta_e$, with each data point being the average result from a particular pair of coins. The values of S' used in determining S'/θ in Figure 3b were obtained from the average apparent sizes of monocularly observed coins obtained under similar conditions in the study by Epstein (1963, Table II) discussed previously. From Figure 3b it is clear that the perceived depth between coins (as measured in the comparison field) and the algebraic difference between values of S'/θ are related ($r = .90$) and, in general, satisfy Equation 2 with $1/K_1 = 1$ (although there is possibly some tendency for the relation to be nonlinear for larger differences in S'/θ). Again, these results can be interpreted either as valid evidence for Equation 2 (and 1) or, as an artifact resulting from the use of a comparison field. In a recent study, Epstein (1964), using a comparison field, had 18 *O*s adjust nonsense-form comparison figures to duplicate the apparent depth between simultaneously presented pairs of nonfamiliar, luminous squares of different retinal sizes presented in a dark experimental field. The radial difference between the adjustments in the comparison field for each pair of squares was considered to be the perceived depth (d') between the pairs of squares. If it is assumed that the perceived size of the squares was constant throughout this experiment, Equation 2 can be tested by plotting the average obtained values of d' against $1/\theta_f - 1/\theta_e$. Using this procedure, it was found that the data do not fit a straight line which passes through the origin as Equation 2 requires. The results from this experiment do not support Equation 2.

Retinal Size Ratios as Size Cues to Relative Distance

Based upon the conclusions that perceptually significant cues (*a*) involve relative rather than absolute retinal size and (*b*) occur between adjacent rather than displaced objects, it has been suggested (Gogel, 1963) that "ratios of θs

rather than θs *per se* [p. 116]" are the retinal stimuli for the relative size cue to relative depth. According to this point of view, two perceived depth intervals d'_{ef} and d'_{gh} would be perceived as equal, using the relative size cue resulting from similar objects, only when the ratios of the retinal sizes between each pair of objects are equal. That is, assuming that $S'_e = S'_f = S'_g = S'_h$,

$$d'_{ef} = d'_{gh} \text{ only when } \theta_e/\theta_f = \theta_g/\theta_h. \quad [4]$$

Equation 4 can be contrasted with Equation 2, with the latter stating that if $S'_e = S'_f = S'_g = S'_h$,

$$d'_{ef} = d'_{gh} \text{ only when } 1/\theta_f - 1/\theta_e = 1/\theta_g - 1/\theta_h. \quad [5]$$

Three of the pairs (e and f) of retinal stimuli involved in the recent study by Epstein (1964) produced a constant ratio of retinal sizes θ_e/θ_f and resulted in a nearly constant value of perceived depth d'_{ef}. It was concluded by Epstein (1964) that the results from this study support the hypothesis that the perceived depth between the pairs of objects was determined by the ratio of their retinal sizes. Within the limitations involved in the use of comparison fields for the measurement of depth between nonfamiliar objects, this conclusion supports Equation 4. An equation relating d' and size cues probably should satisfy at least the following conditions: (a) the requirement that S'/θ is the significant variable in the size cue to relative depth; (b) the requirement that d' is zero, positive, or negative when S'_e/θ_e is equal to, less than, or greater than S'_f/θ_f, respectively; (c) the probability that ratios of θs rather than θs per se are the perceptually significant stimuli; (d) the probability that D' is a negatively accelerated function of S'/θ. An example of one of the simplest equations satisfying these requirements is:

$$d'_{ef} = \frac{S'_e}{T} \ln \frac{S'_f/\theta_f}{S'_e/\theta_e}, \quad [6]$$

where T is an observer constant. Equation 4, for example, can be derived directly from Equation 6. It is obvious, however, that additional research will be required, and preferably by methods other than those using comparison fields, before the psychophysical function of the size cue to relative depth can be specified with confidence.

Concluding Remarks

It is concluded that both perceived size S' and retinal size θ are involved in both the relative (retinal) size cue and the familiar (or assumed) size cue to relative depth. Furthermore, it is asserted that these two factors are the only factors involved in the size cue to relative depth. When the two factors S' and θ are expressed as the ratio S'/θ they can be shown to be applicable to any objects regardless of the shape or complexity of the objects. It is clear that, from this point of view, size cues to relative depth occur, for example, between a rock and a tree as well as between two playing cards or other symmetrical objects. The S'/θ value of each of the two (or more) objects must be considered in predicting the perceived depth which will occur between them. It remains to specify the exact relation which occurs between perceived depth and the relative values of S'/θ. The evidence that the size-distance invariance hypothesis in its usual form can specify this relationship at first appears conclusive. However, much of the evidence for this conclusion was obtained by using comparison fields for the measurement of perceived distance. It is asserted that the use of a comparison field introduces an artifact in the measurement of perceived depth from the size cue. An analogous case would be an attempt to measure the effect of the convergence of the eyes on perceived distance by measuring the perceived position of a binocularly observed, experimental object with respect to a binocularly observed, comparison field. The

experimental object would appear equidistant with that object in the comparison field with respect to which its binocular disparity was zero. Clearly, this result would have no bearing on the possible relation between convergence and perceived distance. Similarly, an experimental object with a particular value of S'/θ will appear equidistant with that object in the comparison field which has the same value of S'/θ. This result has no bearing upon whether familiar size is a cue to perceived absolute distance or, more generally, no bearing upon the validity of the size-distance invariance hypothesis (Equation 1). Equation 1 as it applies to size cues may be correct. But, such a conclusion cannot be made with confidence on the basis of evidence resulting from the use of comparison fields. Furthermore, there is some evidence which indicates that, contrary to Equations 1 (and 2), the perceived distance resulting from size cues is not a linear function of S'/θ. An hypothesis which might be called the ratio hypothesis is tentatively suggested (Equation 6) as meeting certain requirements of the operation of the size cue to relative depth.

If, as seems likely, the perceptual effect of relative values of S'/θ can occur between the experimental and comparison field, even though the two fields are not presented simultaneously, the specification of comparison fields becomes more general. A comparison field is any visual field which O has observed prior to, or simultaneously with, the observation of the experimental object. (It is also possible that a visual field which occurs following the presentation of the experimental object might affect the perceptions associated with the experimental object.) This more general definition of comparison fields calls attention to the possibility that the occurrence of visual fields prior to the experiment can affect the perception of the distance or the size of the experimental object even though no comparison fields are formally presented in the experiment. As mentioned previously, however, it seems reasonable that the magnitude of the interaction between the experimental field and the comparison field (as defined either in the specific or more general sense) would be a decreasing function of the time intervals between their occurrences.

The more general definition of comparison fields can be used in explaining the difficulty involved in investigating the perceived absolute distances of familiar objects when using successive presentations of the experimental objects to the same O. A previously presented experimental object becomes a comparison object for the following experimental presentation. The methodological error involved in attempting to measure the perceived absolute distance of the experimental objects under these conditions is the same as that which occurs with the use of any comparison field. Objects or fields presented previously to (or simultaneously with) an experimental object provide values of S'/θ with respect to which the S'/θ value of the experimental object can be compared. Therefore, in these circumstances, the perceptual localization of an experimental object is likely to be determined by its size cue (S'/θ value) with respect to the comparison field (or the previously presented experimental object) rather than by the size cue with respect to its distance from O.

The importance of S'/θ is not limited to the specification of the size cue to relative depth. It has been demonstrated in several studies (Gogel, 1960, 1964a) that the perceived depth from binocular disparity varies as a function of the S'/θ values in the perceptual vicinity of the objects producing the disparity. Therefore, the concept of S'/θ may be the concept which relates the binocular disparity and size cue systems. If so, it is likely that the prevalent distinction between these two systems as primary and

secondary cues is inappropriate since both cue systems are subject to the operation of a common factor.

REFERENCES

ADELSON, D. The relational influence of size on judgments of distance. *J. gen. Psychol.*, 1963, **69**, 319–333.

BAIRD, J. C. Retinal and assumed size cues as determinants of size and distance perception. *J. exp. Psychol.*, 1963, **66**, 155–162.

CARR, H. A. *An introduction to space perception.* New York: Longmans, Green, 1935.

EPSTEIN, W. The known-size-apparent-distance hypothesis. *Amer. J. Psychol.*, 1961, **74**, 333–346.

EPSTEIN, W. The influence of assumed size on apparent distance. *Amer. J. Psychol.*, 1963, **76**, 257–265.

EPSTEIN, W. Effective difference in influence of relative size on perceived relative distance. *Percept. mot. Skills*, 1964, **18**, 383–384.

EPSTEIN, W., & BARATZ, S. S. Relative size in isolation as a stimulus for relative perceived distance. *J. exp. Psychol.*, 1964, **67**, 507–513.

EPSTEIN, W., PARK, J., & CASEY, A. The current status of the size-distance hypotheses. *Psychol. Bull.*, 1961, **58**, 491–514.

GOGEL, W. C. Perception of the relative distance position of objects as a function of other objects in the field. *J. exp. Psychol.*, 1954, **47**, 335–342.

GOGEL, W. C. Relative visual direction as a factor in relative distance perceptions. *Psychol. Monogr.*, 1956, **70**(11, Whole No. 418). (a)

GOGEL, W. C. The tendency to see objects as equidistant and its inverse relation to lateral separation. *Psychol. Monogr.*, 1956, **70**(4, Whole No. 411). (b)

GOGEL, W. C. Perceived frontal size as a determiner of perceived binocular depth. *J. Psychol.*, 1960, **50**, 119–131.

GOGEL, W. C. Convergence as a cue to absolute distance. *J. Psychol.*, 1961, **52**, 287–301.

GOGEL, W. C. The visual perception of size and distance. *Vis. Res.*, 1963, **3**, 101–120.

GOGEL, W. C. The perception of depth from binocular disparity. *J. exp. Psychol.*, 1964, **67**, 379–386. (a)

GOGEL, W. C. Visual perception of spatial extent. *J. Opt. Soc. Amer.*, 1964, **54**, 411–416. (b)

GOGEL, W. C., & HARKER, G. S. The effectiveness of size cues to relative distance as a function of lateral visual separation. *J. exp. Psychol.*, 1955, **50**, 309–315.

GOGEL, W. C., HARTMAN, B. O., & HARKER, G. S. The retinal size of a familiar object as a determiner of apparent distance. *Psychol. Monogr.*, 1957, **71**(13, Whole No. 442).

HASTORF, A. H. The influence of suggestion on the relationship between stimulus size and perceived distance. *J. Psychol.*, 1950, **29**, 195–217.

HOCHBERG, C. B., & HOCHBERG, J. E. Familiar size and the perception of depth. *J. Psychol.*, 1952, **34**, 107–114.

HOCHBERG, C. B., & HOCHBERG, J. E. Familiar size and subception in perceived depth. *J. Psychol.*, 1953, **36**, 341–345.

HOCHBERG, J. E., & MCALISTER, E. Relative size vs. familar size in the perception of represented depth. *Amer. J. Psychol.*, 1955, **68**, 294-296.

ITTELSON, W. H. The constancies in perceptual theory. *Psychol. Rev.*, 1951, **58**, 285–294. (a)

ITTELSON, W. H. Size as a cue to distance: Radial motion. *Amer. J. Psychol.*, 1951, **64**, 188–202. (b)

ITTELSON, W. H. Size as a cue to distance: Static localization. *Amer. J. Psychol.*, 1951, **64**, 54–67. (c)

ITTELSON, W. H. A note on "familiar size and the perception of depth." *J. Psychol.*, 1953, **35**, 235–240.

ITTELSON, W. H. *Visual space perception.* New York: Springer, 1960.

ITTELSON, W. H., & AMES, A., JR. Accommodation, convergence and their relation to apparent distance. *J. Psychol.*, 1950, **30**, 43–67.

KILPATRICK, F. P. The nature of perception: Some visual demonstrations. In F. P. Kilpatrick (Ed.), *Explorations in transactional psychology.* New York: New York Univer. Press, 1961.

KILPATRICK, F. P., & ITTELSON, W. H. The size-distance invariance hypothesis. *Psychol. Rev.*, 1953, **60**, 223–231.

OVER, R. Size and distance-estimates of a single stimulus under different viewing conditions. *Amer. J. Psychol.*, 1963, **76**, 452–457.

SMITH, W. M. Sensitivity to apparent movement in depth as a function of "property of movement." *J. exp. Psychol.*, 1951, **42**, 143–152.

SMITH, W. M. Past experience and the perception of visual size. *Amer. J. Psychol.*, 1952, **65**, 389–403. (a)

SMITH, W. M. Sensitivity to apparent movement in depth as a function of stimulus dimensionality. *J. exp. Psychol.*, 1952, **43**, 149–155. (b)

SMITH, W. M. Effect of monocular and binocular vision, brightness, and apparent size on the sensitivity to apparent movement in depth. *J. exp. Psychol.*, 1955, **49**, 357–362.

UPDEGRAFF, R. The visual perception of distance in young children and adults: A comparative study. *U. Ia. Stud. Child Welf.*, 1930, **4**(4).

(Received February 11, 1964)

Psychological Bulletin
1964, Vol. 62, No. 4, 236–247

THEORIES OF DREAM FORMATION AND RECENT STUDIES OF SLEEP CONSCIOUSNESS[1]

DAVID FOULKES[2]

University of Chicago

Some of the implications of recent electrophysiological studies of sleep for dream theory, especially the theory of dream formation, are considered. Studies showing a variety of mentation in all stages of sleep fail to confirm Freud's belief that mental activity suddenly attracts consciousness at dream onset. The nature of predream mentation, however, supports Freud's concept of day residues, and there is also evidence to support his position that dream-work processes distort these day residues into sometimes barely recognizable components of bizarre dream episodes. Adler's insistence on the continuity of waking and sleeping thought finds support in the nature and extent of nondreaming mentation in sleep. Both Freud and Adler are challenged on the alleged traumatic affective instigation of dreams; emotional complexes may take advantage of the dreaming state, but they seem neither to precipitate that state nor to determine its initial perceptual-ideational content.

Dement and Kleitman (1957b) reported data confirming earlier and more tentative findings by Aserinsky and Kleitman (1955) on the association of dreaming with the rapid eye movement (REM) periods of sleep. Studies of the undisturbed sleep of normal human subjects (Dement & Kleitman, 1957a) had revealed the cyclic recurrence during sleep of periods of low voltage, random EEG (Stage 1) accompanied by intermittent, conjugate REMs. Dement and Kleitman (1957b) made awakenings during these periods and also during periods of non-REM (NREM) sleep. Their criterion of dream recall was that

> *S*s were considered to have been dreaming only if they could relate a coherent, fairly detailed description of dream content. Assertions that they had dreamed without recall of content, or vague, fragmentary impressions of content, were considered negative [p. 341].

Their results were that dream recall, thus defined, was obtained on 80% of REM-period awakenings but on only 7% of NREM-period awakenings. It was further noted that awakenings made within more than a few minutes of the termination of undisturbed REM periods produced almost no dream recall, a finding taken to indicate that the content of a REM-period dream is swiftly forgotten unless an awakening occurs during the REM period.

The findings of this study have led to the widespread adoption of a new technique of dream retrieval—the interruption of REM periods by experimental awakenings. This technique seems to promise the dream researcher a more representative and complete sampling of the dreaming process than has been accessible previously and to provide an opportunity for the observation of dreams "in progress," that is, they may be interrupted at various points of their development and almost immediate recall obtained. Such a technique, therefore, also seems to present unique op-

[1] The writing of this paper was supported by National Institute of Mental Health Research Grant M-4151.

The author is grateful to Heinz L. Ansbacher, Allan Rechtschaffen, and Gerald Vogel for their critical comments upon its earlier versions.

[2] Now at University of Wyoming.

portunities to test many of the formulations of psychoanalytic dream theory.

One might theorize about many aspects of dreaming. Psychoanalytic dream theorists have been much concerned with the meaning of dream content. They attempt to provide guidelines for the explanation and interpretation of "manifest" dream content, feeling that such content conceals, but may be made to reveal, information about the dreamer and his conception of his world. Dream theories are also concerned with the functional significance of the dreaming process, with statements as to why we dream at all, what psychological functions are served by dreaming, and at least inferentially, what might happen to us were we unable to dream. Comprehensive dream theories have also been concerned with describing dream formation and development, that is, they purport to describe the processes of construction through which the dream becomes an organized psychological event.

In this paper, we shall consider the implications of recent electrophysiological dream studies for the portions of two major dream theories, those of Freud (1956) and Adler (1958), which deal with dream formation and dream development. It is proposed, first, to present an outline of Freudian and Adlerian theories of dream formation and dream development, and then to examine these theories in the light of recent experimental data. In the presentation of Freudian theory, concentration will be focused upon those formulations which lend themselves to some kind of reasonably direct empirical verification.

Two Theories of Dream Formation

The central propositions of Freud's (1956) theory of dream construction or dream formation are:

1. Dream formation and early phases of the organization of the dream will not be represented in the conscious experience of the dreamer. The typical condition of sleep is a state of unconsciousness, a state in which no mental activity is available to personal awareness or to report to others. In discussing the "psychical conditions during the period of sleep which precedes dreams," that is, the period in which the dream is being "formed," Freud (1956) says "that we are dealing with an *unconscious* process of thought . . . [p. 281]." Much of the groundwork which has gone into the construction of the dream as an intelligible perceptual event takes place before the dream, in Freud's term, "attracts" consciousness to itself. Freud suggests that the dream "is like a firework, which takes hours to prepare but goes off in a moment." This preparation is achieved outside the boundaries of consciousness. It is only when the intensity of these unconscious processes becomes sufficient to arouse mechanisms of consciousness, or when, "just before waking, attention becomes more mobile and comes to meet it," that we experience the dream as a conscious event (p. 576).

2. The unconscious process which instigates dream formation is affective in character, and in particular, it is some derivative of the primary motives of sex or destructive hostility. In sleep, unexpressed and unexpressible sexual drives or hostility are freed from external monitoring, and also from the kinds of internal monitoring which require a high degree of cerebral vigilance. These impulses, long active, now press for some kind of expression. The dream provides the occasion for such expression.

In its final form, the dream is a distorted and symbolic rather than a manifest, direct expression of the impulses

which instigate it. This reflects the fact that the forms of waking cortical inhibition exercised over primitive emotional processes are not entirely lacking during sleep. The dream is a fruit of compromise: on the one hand, it does provide an outlet for the expression of primitive emotional impulses; on the other hand, this expression is not (generally) so blatant that the dreamer's critical and inhibitory faculties will be extremely offended by its content. The manifest content of the dream, then, is not a direct reflection of the latent instigating "dream thoughts"; it is because of this disjunction that the manifest dream requires "interpretation."

3. If repressed affect provides the energy for dream construction, perceptual-memorial events provide the raw material. The impulses find their way to expression along sensory, rather than motor, pathways. The dreamer does not, with a patterned sequence of motor activities, inflict harm upon another person nor even go through all the appropriate motions (but, see Wolpert, 1960); he may, however, visually hallucinate that he is harming someone. The kind of thinking in which sensory hallucination of a goal serves as a substitute for its actual achievement Freud calls *primary process* thinking. Dreams represent the most acute manifestation in the adult human of this developmentally primitive, perceptual-hallucinatory, mode of thought.

The raw materials used in dream construction must, therefore, be traces of prior perceptual experiences of the dreamer. In particular, Freud (1956) feels, very recent experiences, experiences from the day immediately preceding the dream, serve as basic elements in dream construction. These elements Freud calls *day residues*. The day residues will often consist of memories of what appear to be rather inconsequential happenings. But it is not what these memories are that is important; it is the fact that they can represent, or serve as screens for, the repressed impulses. The more inconsequential the day residues are, the better they may fulfill their function of disguising socially unacceptable impulses.

But Freud (1956) notes that dreams include much "infantile" perceptual material as well as the "recent and indifferent" material represented in day residues. Dreams, in fact,

> can select their material from any part of the dreamer's life, provided only that there is a train of thought linking the experience of the dream-day (the "recent" impressions) with the earlier ones [p. 169].

It is through their association, then, with day-residue material that older sensory memories are contacted, and, in turn, it is the ability to contact such memories that plays a role in the determination of which day residues shall figure most prominently in the formation of the dream.

4. The day residues, and their associated infantile memories, are the basic elements of the dream, but it would be misleading to conceive the dream as an orderly and logical sequence of such memories. Freud speaks, in particular, of day residues as being "worked over" so that the ultimate dream product is a complicated and bizarre patterning of the original elements. In fact, the elements may be so transformed in the dream that it will be difficult to establish precisely what they are. These transformations, which highly complicate the task of dream interpretation, appear for several reasons: so that the repressed impulses may be adequately expressed, that is, for purposes of dramatic representation; so that the repressed impulses may be successfully disguised, that is, for purposes of evading censorship; and so that the pres-

entation will be as economical as possible, that is, for purposes of the conservation of psychic energy. These are the functions served by a series of processes to which Freud applies the term *dream work*. These processes include: condensation, displacement, and symbolization. The operations of these processes make the dream progressively less intelligible to waking consciousness and progressively detached from its moorings in unretouched sensory memory.

Let us now examine, briefly, one other theory of dream formation, that of Alfred Adler (1958). While Adler's theory of dreaming is by no means as comprehensive or detailed as is Freud's, it has gained much significance because it has formed a large part of the conceptual foundation of several recent and highly influential dream theories, those of Erich Fromm (1957) and Calvin Hall (1959). It is possible, moreover, to find some assertions about the processes of dream formation and dream development in Adler's (1958) own writings and in a recent restatement of the Adlerian position by Ullman (1962), assertions at considerable variance from those made by Freud.

1. Sleeping and waking thought are not totally incompatible with one another; we must recognize the essential continuity of all forms of thought. In particular, Adler objects to Freud's conception of dream thought as determined by a mechanism relatively inoperative in waking thought—the unconscious—and to the distinction of primary process and reality-centered thinking, with the former characterizing sleep and the latter wakefulness. There is, to be sure, some difference between dream thought and waking thought, but it is a relative, rather than an absolute, one. The dreamer maintains fewer relations with reality. Yet there is no complete break with reality; he is still in contact with it.

2. The instigation of dreams is not always, or even often, due to sexual or hostile motives, any more than waking thought is generally dominated by such motives. Again, Adler insists that the dream cannot be a contradiction of waking life; it is always consonant with one's waking style of life.

In common with Freud, Adler feels that we dream when we are troubled by something. We dream only when unresolved problems from waking life, which Ullman (1962, p. 20) characterizes as the "sore spots" of one's existence, press upon us during sleep. It is the task of the dream to meet and to try to solve such problems. The troublesome "something" which instigates the dream is, then, a problem from conscious experience, not a problem which has been repressed and of whose existence the waking organism is totally unaware. Freud (1956), on the other hand, relegates such problems to a "secondary position [p. 554]" with respect to dream formation.

3. The raw material of the manifest dream content comes from memories of prior perceptual experience, particularly from day residues. But, in Adlerian theory, the day residues are important in themselves, as representations of waking concerns; they are not important simply as "screens" for something else. Adler does not share Freud's (1956) feeling that day residues are "some cheap material [p. 237]" of little significance in determining the direction of the dream.

4. In the dream, these residues are transformed or worked over until the ultimate product becomes a deceitful working through of a waking problem. Adler (1958) speaks of the dream as being constructed "to fool us [p. 107]." As Ullman (1962, pp. 23–24) has

pointed out, there is some contradiction here of Adler's basic theme of the continuity of waking and sleeping thought. The mechanisms of distortion which Adler seems to recognize include condensation, displacement of emphasis, and symbolization. There seems to be a greater stress in Adler than in Freud, however, upon the expressive nature of symbolic representation. In dreaming, we make use of those images and incidents which best agree with our style of life and which best express the present problem. This relatively strong emphasis on the symbol that expresses rather than the symbol which disguises is consistent with the notion of a continuity between waking and sleeping thought and becomes the keystone of Fromm's (1957) and Hall's (1959) theories of dreaming.

Electrophysiological Studies and Dream-Formation Theory

It is apparent from the Dement and Kleitman data that both Freud and Adler misjudged the conditions under which dreaming occurs. Both seem to feel that we dream as it is needed, that is, in proportion to the number and intensity of our personal problems. But the facts seem to be that most adult human beings spend approximately 20% (e.g., Dement & Kleitman's figure, 1957a, is 18%) of their sleep in REM periods, periods which produce the kinds of mental content with which both theories are concerned. These periods recur in all individuals studied in a highly predictable cyclic fashion. Individual variations from the 20% figure have been minimal and there is little evidence that they are related to the nature or extent of personal problems. We seem to dream as physiological cycling indicates, not as relative degree of latent psychological trauma dictates. It is still possible, of course, that emotional complexes take advantage of the dreaming process, but apparently they do not, at will, precipitate that process.

The Dement and Kleitman (1957b) recall data were collected well after dream onset in the case of REM-period awakenings and other than immediately before REM onset in the case of NREM awakenings. For these reasons they provide an insufficient base for further evaluation of theories concerning events occurring at dream onset. Several more recent studies, however, have arrived at basically similar data which seem highly relevant to the characterization of the processes of dream formation and dream development.

Studies of Pre-REM Mentation

Foulkes (1962) conceived his study as an objective study of dream formation. Eight subjects were run for 7 nights each. Awakenings were made at various points in the sleep cycle clustered around the onset of REM periods, those portions of the sleep cycle in which, according to Dement and Kleitman (1957b), dreams would be occurring. It was reasoned that such awakenings might provide an empirical basis for theorization about dream onset and the characteristic sequence of dream development. Foulkes found, however, that subjects produced reports of mental phenomena from NREM sleep almost as often as they did from REM sleep. Apparently no point of absolute dream onset exists, in the sense that there is no point in the sleep cycle at which consciousness suddenly appears. It seems to be there all along.

There was, however, some indication of progressive changes in the quality of sleep consciousness around the point of REM onset. More generally, content from pre-REM and REM stages of sleep differed systematically along a number of dimensions. As assessed by

direct questioning of the subjects during the night and by responses to rating forms taken on the following morning, reports from the pre-REM periods were less likely than REM reports to be labeled as "dreams," more likely to be called "thoughts"; less likely than REM reports to be vivid or highly elaborated; and more likely than REM reports to be intimately associated with recent and everydayish activities of the subject, sometimes to the point of consisting purely of memories of such events.

The nature of these differences may be appreciated more fully perhaps with several examples of REM and NREM reports. The following reports were obtained from an adult male subject who was employed by the Internal Revenue Service:

1. He asked an acquaintance at work for a hammer, so that he could fix something in his apartment. (NREM)

2. He was thinking of a point made in his tax class, that you have to provide over half of a person's support to claim him as a dependent. (NREM)

3. He received a phone call in the middle of the night from a girl identifying herself as from the University of Chicago. She said that it was time for his "35-day evaluation." He chided her for calling so late at night. She replied that it was the only time they could get him in. (REM, 3 minutes after onset of eye movements)

The first two reports, which are typical of this subject's NREM reports, have an everydayish quality which the third one lacks. In commenting on the 35-day evaluation by the University of Chicago, with which his only connection was his service in the dream study, the subject noted that he was in a 90-day probationary period in his new job. He was to receive $35.00 for his services as a dream subject. It seems, then, that experimental and work experiences have been fused in the REM period dream so that neither experience is portrayed with complete accuracy. In the second report, however, the subject is reexperiencing, in a completely undistorted manner, the recent experience of considering one of the details of his new job. The purely conceptual quality of this report was a fairly typical NREM characteristic for three of the eight subjects in this study.

The other subjects generally had somewhat more "dreamlike" NREM reports. The following examples come from such a subject, an undergraduate major in English literature.

1. He pictures Anna Karenina. She is sitting at a table, then gets up, turns to the left, and walks away. (NREM)

2. He is in a sleep laboratory, filling out a pencil and paper form. Someone passes by, commenting that the task is a stupid one. (NREM)

3. In the first scene, he is standing on a street corner, holding his bicycle, and talking to someone about a girl who wanted to be a striptease dancer.

In the second scene, he is in a doctor's study with two women and the doctor. They are discussing two books. The heroine in the first book was a striptease dancer, but is no longer this, but a nurse. The women are discussing how much hardship she has as a nurse. A discussion then ensues of a second book, by John Steinbeck, in which the main character, also a nurse, did not, apparently, endure similar hardship. The women discuss this avidly, as if they were going to go into "this sort of thing." (REM, 3 minutes after onset of eye movements)

The first two reports of this series seem to be tied much more clearly, and with much less distortion, to recent manifest behaviors or concerns of the subject. In the first, he brings to life a character of a novel which he is reading; in the second, he takes notice of the questionnaire which he must fill out each evening at the laboratory, and makes, through a vaguely identified figure, a comment as to how he views this labor. Although these NREM reports are more detailed and dreamlike than those of

the previous subject, there is still a striking difference with REM content in terms of distortion and elaboration.

Confirmation of the reliability of the differences found in the Foulkes (1962) study between REM and NREM material has been obtained by Rechtschaffen, Verdone, and Wheaton (1963). These authors studied the REM and NREM mentation of 17 normal subjects for a total of 30 nights and report that:

> In every instance where the variable being measured was comparable, i.e., recall, thinking-dreaming, vividness, visual-conceptual, volitional control, plausibility, time referent, and emotionality, Foulkes' results were in the same direction as ours. The two studies were done entirely independently, and neither investigator knew the variables the other was studying [pp. 410–411].

In the course of a study of the effects of presleep stimuli on dream content (24 subjects, 2 nights each), Foulkes and Rechtschaffen (1964) have reported still further confirmation of the reliability of the differences noted above between REM and NREM content.

But can the NREM reports of these three studies be accepted as valid indicators of mental processes occurring during NREM sleep? Can we be sure that these reports are not awakening artifacts, representing material experienced after the onset of the awakening stimulus? Can we be sure that these reports are not recollections of material experienced in previous REM periods? The available evidence, discussed elsewhere (Foulkes, 1962; Rechtschaffen, Verdone, & Wheaton, 1963) would seem to indicate that most NREM content is just what it appears to be: it represents experiences which occur during NREM sleep. The failure of Dement and Kleitman (1957b) to uncover this kind of material, which failure is responsible for raising the question of validity with respect to later findings on NREM recall, appears to be attributable to their very stringent criterion of recall, quoted above, and to several features of their awakening procedure which probably served to depress recall outside REM periods (Foulkes, 1962).

Pre-REM Mentation and Dream Formation Theory

What, then, do these apparently consistent and valid findings on mentation occurring outside REM periods suggest about the validity of the two theories of dream formation which we are examining?

1. Freud's characterization of the sleep in which full-fledged dreams are not occurring as periods of unconsciousness seems to be inaccurate. The typically bizarre and elaborate REM-period dream does not burst like a sudden firework against a background of complete darkness; it develops in a context of already ongoing mental activity.

Parenthetically, we may inquire whether someone of Freud's stature could have been totally unaware of the existence of the everydayish thinking which seems to take place during much of the sleep cycle. As we might suspect, the answer is no. In Freud's (1959) paper "Dreams and Telepathy," he notes that there are mental events during sleep without condensation, distortion, dramatization, and wish fulfillment. These unaltered repetitions of actual daily experiences he calls *night phantasies* (p. 421). That the admission of night phantasies has not played any major role in Freud's theory of the dream process may be seen, however, in their treatment in *The Interpretation of Dreams* (Freud, 1956). There (p. 331), Freud indicates that he once considered the possibility of such a class of sleep mentation, but later dropped that category. The net impact of his

total theoretical position, therefore, has been that we experience alternating periods of unconsciousness and consciousness during sleep, consciousness appearing when the dream is sufficiently well developed to "attract" it. This now seems to be a misleading characterization of sleep consciousness.

2. There is a strong resemblance between the NREM content which precedes dream periods and the Freudian concept of day residues from which dreams are presumed to develop. Insofar as the day-residue material reveals some of the "background thoughts [Freud, 1956, p. 103]" of the dream, collection of NREM material could facilitate dream interpretation in a more direct manner, perhaps, than even the waking free associations upon which psychoanalytic interpretations have heretofore depended so heavily. The whole body of NREM material occurring between REM-period dreams might be considered the conscious

> exploring of one path and another, a swinging of the excitation now this way and now that, until at last it accumulates in the direction that is most opportune and one particular grouping becomes the permanent one [Freud, 1956, p. 576].

3. The rather dramatic shifts from plausible content to implausible content, from the everydayish to the bizarre, which attend REM onset suggest the engagement at this point of the sleep cycle of processes much like those which Freud calls the dream work. For example, 23% of 26 content-producing reports elicited by Foulkes (1962) from ascending EEG Stage 2 (Dement & Kleitman, 1957a), that portion of the sleep cycle immediately preceding REM periods, were undisguised memories or re-creations of recent events in the dreamer's life, while not a single one of 31 reports taken from awakenings made within 4–60 seconds of REM onset were of a comparable nature.

Moreover, Verdone (1963) has shown that longer REM periods produce reports judged by subjects as more vivid and emotional, and Foulkes' (1960) data reveal that reports from longer REM periods (9–24 minutes of REMs) produced greater mean values in subject ratings for activity, emotionality, anxiety, unpleasantness, frighteningness, dramatic quality, violence-hostility, and distortion than did reports elicited from within 4–60 seconds of REM onset. Whatever the processes are which become active at REM onset, they become increasingly predominant as the REM period progresses. And the highly organized (if often bizarre) drama they generally create supports Freud's attribution of dream distortion to active processes, to motivated condensation, displacement, and symbolization, rather than a conception that dream distortion is a symptom of general mental disorganization due to a sleepy, hence imperfectly functioning, cerebral cortex.

4. Strictly speaking, the fact that day-residue content is often experienced during pre-REM period EEG Stage 2 and that dreamlike experiences occur in that stage of the sleep cycle which immediately follows, namely, EEG Stage 1 REM periods, does not establish that the first kind of material is dynamically transformed to produce the second kind of material. Is there any evidence that the day-residue content of Stage 2 forms the dynamic basis of the dream, that such material is "worked over" as the REM period commences? The following example seems to suggest that this may be the case:

> I was in the library and I was filing cards, and I came to some letter between "a" and "c." I was filing some, I think it was Burma, some country, and just as I put that in, there

was this scene of some woman, who was sent to look for a little girl who was lost, and she was sent to Burma. They thought the little girl was going there, for some reason. This was sort of like a dramatization of what I was doing. I mean I was filing, and then this scene took place, right at the same time. In the setting it was sort of like you'd imagine it, but I had the feeling it was really happening.

In this dream, obtained 26 seconds after REM onset, a scene from the subject's daily work experience—she had a filing job in the university library—led to an elaboration which was far removed from her everyday experience. The daily-work element (filing), typical in Stage 2 content, preceded the unusual and somewhat unrealistic element (Burma), typical in REM-period content. The physiological recording shows a recent progression from Stage 2 to a REM period. This one case is not, of course, in itself conclusive, but it does suggest that REM-period dreams may begin with the working over of day residues of a sort most often experienced during NREM Stage 2.

Rechtschaffen, Vogel, and Shaikun (1963) have recently demonstrated, moreover, that discrete manifest elements and themes found in NREM reports are sometimes repeated in subsequent REM reports from the same night. Such findings led them to conclude that:

On those nights when themes and images persist through both NREM and REM periods, the dreams do not arise *sui generis* as psychologically isolated mental productions, but emerge as the most vivid and memorable part of a larger fabric of interwoven mental activity during sleep [p. 546].

5. Adler's assertion of a basic similarity of waking and sleeping thought finds support, at least when we consider NREM-period sleeping thought. Rechtschaffen, Verdone, and Wheaton (1963) have likened the content of NREM mentation to background mentation in waking experience. It is, in Stekel's (1951) image, a muted accompaniment rather than a melody. In NREM sleep, the melody, or directive character of thought, seems less prominent and we become aware of this background thought. Such an awareness is difficult to achieve in waking life; it is as if we might have direct and immediate access to all those irrelevant things which passed along the borders of consciousness while we were more or less attentively engaged in some line of directed thought. Conditions of sensory deprivation or sensory monotony (Fiske, 1961) perhaps give us the best possible vantage point in waking life for viewing such phenomena. While the identification of NREM mentation with fringe-type waking mentation may not be an apt characterization of all NREM mentation, it does underline the degree of continuity of NREM mentation and at least some forms of waking experience.

In the context of Freudian theory, what we do not find in pre-REM sleep mentation is worthy of note. We do not find a seething, libidinous turmoil, sharply at variance with waking thought. Rather we see a generally "relaxed" kind of thinking and imaging whose prominence becomes possible in the absence of a strong external focus or challenge for thought processes and whose nature is essentially similar to certain classes of incidental waking mentation.

6. But, if we find little evidence of libido or hostility in the pre-REM thought from which dreams may be presumed to develop, neither do we find much evidence for the Adlerian assertion that representations of pressing personal problems of a more general character are active at, and responsible for, dream formation. Awakenings made at various points leading into REM-pe-

riod onset and during the early seconds and minutes of REM periods do not corroborate, at least in any obvious manner, the position that dreams begin with affective or ideational "sore spots." Transformations are noted in the quality of mental experience: from conceptualizing or vaguely perceiving to vivid visual imaging, from kaleidoscopic flux to continuity and integration, and from plausible content to distorted and bizarre content. However, these changes do not seem to be associated with any perception of particularly pressing personal problems which cry for some resolution or exploration. Rather than being sources of personal anxiety or insecurity, NREM contents of consciousness generally seem quite unconnected with basic psychodynamic concerns of the dreamer.

Because this last point is a particularly crucial one for both theories, we will do well to consider it in further detail. Will the dream theorist be impressed with evidence that the class of events-in-consciousness at or around REM-period-dream onset fails to meet his specifications of dream-instigating events? Most likely not. He might comment that the analysis has, up to this point, relied too heavily upon appearances, upon manifest consciousness. It might be suggested that the everydayish, relaxed kinds of mentation which precede REM-period dreaming serve as screens for affective elements, based in infantile experience, which present a threat, or pose a challenge, to the dreamer. Such a contention raises a basic methodological question, however: since neither physiological nor experiential evidence seems to support the hypothesis that dream periods develop from or commence with such elements, with what kind of evidence does the dream theorist propose to confirm his theory of the affective instigation of dreams? At present, the assumption that pre-REM-period contents are screens for something else seems entirely gratuitous. What we seem to have in NREM-period mentation is one level of cognitive functioning, a level which should at least tentatively be accepted for what it is, rather than be immediately interpreted as a disguised form of something else.

At this point we should, perhaps, consider the observational base upon which both Freudian and Adlerian theory rest, clinical dream collection. Ian Oswald's (1962) comments upon the adequacy of clinical dream collection as a basis for characterization of the dream process cut right to the heart of the matter. He notes that the patient probably describes only a very small fraction of his total dreams to his therapist, and that his "recall" of these dreams is more a construction than a reconstruction. This constructed, waking fantasy-material may tell the clinician a great deal about the patient, but "is not to be relied upon as evidence of what really happens during dream periods [p. 143]."

Yet both Freud and Adler, in the absence of systematic observations made at various phases of the (REM) dream process, have given us theories of that process. Quite clearly, what they have done is to read back into the dreaming process those events which it seems must have transpired in its early stages, given a certain dream outcome and/or a certain functional interpretation of that outcome. Now no one who has had experience in collecting, examining, or interpreting dreams can doubt that dreams often do express a person's basic feelings and indicate his problems. And yet, the pre-REM and early-REM recall data now available suggest that we err in assuming that since this may be the functional significance of the dream,

the dream therefore must have started with the expression of such feelings or the posing of such problems. Rather it seems as if the dream allows the dreamer, eventually rather than immediately, to express himself in a rather profound way; it is not that the dream, by posing a basic challenge at its very onset, forces him to do this.

It is being suggested, then, that both Freudian and Adlerian theory are incorrect in their characterization of the dream process, that is, dream formation and dream development. But what, we may now ask, are the implications of this position for those portions of their dream theories which deal with dream meaning or the functional significance of dreaming? We have already noted that Freud's and Adler's theories of the dream process are generalizations from, rather than the inductive basis of, their theories of dream interpretation. This suggests that dream process data can confirm or disconfirm a particular theory of dream meaning only to the extent that that theory is compatible with one, and only one, characterization of the dream process. Dream theorists have tended to assume such a necessary correspondence between interpretation theory and process theory; that is, they have generalized in a relatively direct fashion from what the dream is to how it must have started. But this overlooks the possibility that how a dream begins and what a dream becomes may be two entirely different matters. Both Freudian and Adlerian theories of dream meaning and function, therefore, may be compatible with the description of the dream process provided by recent electrophysiological research. To reject Freud's dream-process theory is not necessarily to reject his position that dreams serve the function of the fulfillment of repressed wishes. To reject Adler's dream-process theory is not necessarily to reject his position that dreams are attempts at problem solving which are consistent with the dreamer's life style.

Although the present data on the dream process are ambiguous with respect to theories of dream meaning, it is likely that further studies of the content of REM- and NREM-period dreams will provide some basis for choice among these theories. Dement's research on dream deprivation (1960) has already indicated the usefulness of electrophysiological studies in evaluating theories of the functional significance of the dream.

In conclusion, it is hoped that this discussion has indicated some of the possibilities which the Aserinsky, Kleitman, and Dement dream-collection technique offers with respect to resolving or clarifying some of the major issues in classical, mainly psychoanalytic, dream theory. But the major contribution of this technique may well be the generation of altogether new theories of dreaming, theories solidly based upon experimentally derived dream-process data, and theories which articulate with modern neurophysiological evidence (e.g., Jouvet, 1962) upon the nature of the sleep phase in which dreaming occurs.

REFERENCES

Adler, A. *What life should mean to you.* (Orig. publ. 1931) New York: Capricorn, 1958.

Aserinsky, E., & Kleitman, N. Two types of ocular motility occurring in sleep. *J. appl. Physiol.,* 1955, **8**, 1–10.

Dement, W. The effect of dream deprivation. *Science,* 1960, **131**, 1705–1707.

Dement, W., & Kleitman, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *EEG clin. Neurophysiol.,* 1957, **9**, 673–690. (a)

Dement, W., & Kleitman, N. The relation of eye movements during sleep to dream activity: An objective method for the study

of dreaming. *J. exp. Psychol.*, 1957, **53**, 339–346. (b)

Fiske, D. W. Effects of monotonous and restricted stimulation. In D. W. Fiske & S. R. Maddi (Eds.), *Functions of varied experience.* Homewood, Ill.: Dorsey Press, 1961. Pp. 106–144.

Foulkes, W. D. Dream reports from different stages of sleep. Unpublished doctoral dissertation, University of Chicago, 1960.

Foulkes, W. D. Dream reports from different stages of sleep. *J. abnorm. soc. Psychol.*, 1962, **65**, 14–25.

Foulkes, D., & Rechtschaffen, A. Presleep determinants of dream content: The effects of two films. Paper read at Midwestern Psychological Association, St. Louis, May 1964.

Freud, S. *The interpretation of dreams.* (Orig. publ. 1900) New York: Basic Books, 1956.

Freud, S. Dreams and telepathy. In, *Collected papers.* Vol. 4. (Orig. publ. 1922) New York: Basic Books, 1959. Pp. 408–435.

Fromm, E. *The forgotten language.* (Orig. publ. 1951) New York: Grove Press, 1957.

Hall, C. S. *The meaning of dreams.* (Orig. publ. 1953) New York: Dell, 1959.

Jouvet, M. Recherches sur les structures nerveuses et les mecanismes responsables des differentes phases du sommeil physiologique. *Arch. Ital. Biol.*, 1962, **100**, 125–206.

Oswald, I. *Sleeping and waking: Physiology and psychology.* Amsterdam, The Netherlands: Elsevier, 1962.

Rechtschaffen, A., Verdone, P., & Wheaton, Joy. Reports of mental activity during sleep. *Canad. Psychiat. Ass. J.*, 1963, **8**, 409–414.

Rechtschaffen, A., Vogel, G., & Shaikun, G. Interrelatedness of mental activity during sleep. *Arch. gen. Psychiat.*, 1963, **9**, 536–547.

Stekel, W. The polyphony of thought. In D. Rapaport (Ed.), *Organization and pathology of thought.* (Orig. publ. 1924) New York: Columbia University, Bureau of Publications, 1951. Pp. 311–314.

Ullman, M. Dreaming, life-style, and physiology: A comment on Adler's view of the dream. *J. indiv. Psychol.*, 1962, **18**, 18–25.

Verdone, P. P. Variables related to the temporal reference of manifest dream content. Unpublished doctoral dissertation, University of Chicago, 1963.

Wolpert, E. A. Studies in psychophysiology of dreams: II. An electromyographic study of dreaming. *Arch. gen. Psychiat.*, 1960, **2**, 231–241.

(Received November 18, 1963)

Psychological Bulletin
1964, Vol. 62, No. 4, 248–256

PREDICTING GROUP TASK EFFECTIVENESS FROM MEMBER CHARACTERISTICS

RICHARD HESLIN [1]

University of Colorado

Studies relating premeasures of member characteristics to small-group task performance are reviewed. The individual traits used to predict performance are grouped into 6 categories: ability, adjustment, extraversion, dominance, authoritarianism, and "other characteristics." 2 member characteristics have been used as predictors in a sufficient number of studies to permit a tentative conclusion, namely, that member ability (both specific and general) and member adjustment are significantly related to group effectiveness. It is less clear however, whether extraversion, dominance, and authoritarianism would be useful for predicting group task performance. Both the requirements of the situation and social structural constraints must be taken into account when predicting group effectiveness from member traits. Specifically the type of group task and the organization of the group are seen to have a decided influence on the relationship.

The major determinants of group effectiveness have traditionally been sought from among situational demands and factors stemming from the prevailing social system both within and outside the particular group under study. Although the personal characteristics of individual members have received much less attention, it is becoming increasingly apparent that the external and social variables interact with personality characteristics of the participants, and that the group product is at least partly determined by factors the members bring to the task. This review is aimed at examining to what extent group effectiveness is determined by personality characteristics of the members.

This question has significant practical utility since many organizations measure various personality characteristics of their new members and are interested in the performance of their committees and work teams. If they can use some of the information they have available on the personality of their members to construct more effective groups, a large stride toward rational and efficient use of people will have been made.

Small group member characteristics are usually measured either by rating or categorizing individual behavior in ongoing groups, or by completing questionnaires on group-tied characteristics such as satisfaction with the meeting. All of these measures require that the members either are interacting or have already interacted. The question toward which this paper is directed is whether a measure of individual members taken before the group is formed can be used to predict its task effectiveness.

The studies reviewed are grouped under the following rubrics that typically appear (as factors) in factor analytic studies of interpersonal behavior: ability, adjustment, extraversion, dominance, authoritarianism, and other characteristics.

Other characteristics was included to group a number of studies that meet

[1] I would like to thank O. J. Harvey, Richard deCharms, and Keith Davis for their critical reading of the manuscript.

the criteria for inclusion in the review but the personality variables of which do not belong under one of the other five categories. It includes such things as general personality profile, self-perception, tendency to differentiate people sharply, and the desire to be close and intimate with others.

The dependent variable in this review is group performance. In the kinds of studies reviewed in this paper, situational demands are so strong that member definition of group goals is not as necessary as in less-structured small groups. Therefore, definition of the task and evaluation of the final product by an agency outside the group has been taken as a valid measure of the performance of the group.

To those interested in the application of research, the performance of a small group is more central to their interests than its internal workings. But the relation between predictor and performance is often understandable only in terms of some mediating variable. Therefore, two analyses of the problem will be made. The simple relation between group performance and each personality variable will be reviewed in detail, and then the influence of situational and social variables will be noted. Let us turn now to the analysis of the personality predictor variables.

Ability (General)

Leaders of effective army squads have a significantly higher mean intelligence score than leaders of ineffective squads (Greer, Galanter, & Nordlie, 1954), effectiveness here defined as squad performance as rated by judges using detailed performance rating forms. (The studies of army-squad effectiveness usually use a small scale, standardized field maneuver of 6–8 hours with umpires rating the performance of the groups.) The relation between leader intelligence and performance is supported by Havron and McGrath (1961), who found that the correlation between unit effectiveness and leader intelligence ranged between .35 and .50 in their series of studies. Leader intelligence, with job knowledge, shared the position of the best leader trait for predicting group effectiveness.

The relationship of intelligence to group performance is susceptible to influence from other variables. For example, it appears that the organization of the members can modify it. Permissive group organization can facilitate problem-solving performance of groups composed of high-intelligence members but inhibit the performance of groups composed of low-intelligence members, while directive group leadership seems to result in the opposite relation (Calvin, Hoffman, & Harden, 1957). Although only one of the six relationships was significant, beyond the .05 level, the results were all in the same direction.

An early study by Gurnee (1937) compared 42 individuals with 12 groups on a more limited and specific task, learning the correct pathway of a maze. The groups proceeded by plurality vote, voting by acclamation, with a show of hands in case of a tie. Results of interest here were that the number of errors in learning the maze were negatively related to college grades ($r = -.54$) and the chemistry aptitude scores ($r = -.68$) of the group members.

A series of three studies result in a statement (in the third study) on the relation between group performance and the variability of intelligence among the members. First Pelz (1956) reported a study of factors influencing the performance of scientists in a research organization. He found that frequent contact with other scientists was positively related to an individual's performance level only if the contacts were

with scientists unlike him in kinds of motivation and experience. An analogous phenomenon was discovered under more controlled conditions by Hoffman (1959) and by Hoffman and Maier (1961). Only the first study will be discussed. Hoffman administered the Guilford-Zimmerman Temperament Survey to subjects and later formed them into four-member groups on the basis of their scores. Some groups were composed of members whose personality profiles were very similar, others of members whose personality profiles were quite different. The heterogeneous groups produced significantly superior solutions to the multisolution, mined-road problem ($p < .05$) and showed a tendency to produce more inventive solutions to a human-relations problem ($\chi^2 = 2.88$, not significant). Hoffman (1959) stated "Pelz's findings suggest that the results reported in the present study are probably generalizable well beyond the limited population of college students who supplied the data [p. 31]." However, according to Shaw (1960), the superiority of heterogeneous groups is not generalizable to all situations and specifically not to member intelligence in communication-net situations. In two studies, he found group effectiveness was unrelated to group heterogeneity in intelligence.

Shaw's use of the communication-net design reduces the generality of his criticism since communication among the members in this situation is so artificially restrained. Furthermore, his group tasks did not unduly tax the intelligence of his sample (university students). Once a subject had received the necessary information from the other members, he had to perform a series of computations requiring about a fifth-grade knowledge of arithmetic.

The relationship between members' general ability and group performance is not univocal. For example, Spector and Suttell (1957) found no significant relation between average intelligence of three-man groups and performance. However, since (*a*) the performance measure was taken only on the last of four tasks, and since (*b*) the manipulation of leadership style, a second experimental variable, did not produce differences in performance either, it is not clear whether the measures were too gross to expose differences or whether there were no differences to measure.

In general, group task performance was related positively to general ability in three studies and was unrelated in one study; in groups with more formal structure the leader's intelligence increased in importance as a factor influencing performance. In no study was a negative relationship found between the two variables.

Ability (Specific)

McKeachie (1954), in his review of the literature on the success of student-centered methods of instruction, concluded that the success of this method of teaching may well depend on whether or not the students possess the specific skills necessary to move the group toward its goal. Group performance is limited by the members' ability both on the task and in organizing themselves effectively.

An experiment to measure the relation between specific ability and group performance was conducted using the Purdue Pegboard Test (Comrey, 1953). Each subject was first tested on his individual pegboard and later participated as a member of a group performing a similar task. The "group" was a dyad working on a common pegboard; that is, Member A placed a peg, B placed a washer on the peg, A covered the washer with a collar, and B placed a final washer on the assembly while

simultaneously inserting a new peg diagonally across from the old assembly. Thus they zig-zagged down the board alternating their activities as they went. The performance of the group was related to the individual performance of the less-able member, .59, and the more able member, .56. Using the performance of both members as predictor yielded a multiple correlation of .66. If Comrey had measured for test-retest instead of split-half reliability, he could have had an index of the effect of such things as change of set and practice during the time interval between the individual and group measures. The difference between this reliability coefficient and the correlation between individual and group conditions would then indicate the influence of group-related factors. Since this was not done, the above correlations reflect the influence of both time- and group-related factors.

Another investigator whose individual and group measures were similar to those of Comrey was Rohde (1958). Subjects were pretested on a maze task before being assigned a similar problem in three-man groups consisting of one appointed leader and two followers. Group performance scores were significantly related to the individual scores of the most qualified members ($r =$ .63), all individuals ($r = .63$), and followers ($r = .60$), but not with the individual scores of the least qualified subjects ($r = .39$). Regarding the last relation, when a subject knows the correct turn he will show more self-confidence and be more forceful in his arguments than when he is unsure of it; thus the more able subjects will have more influence on the group product than the less qualified. This is in contrast with Comrey's (1953) findings with dyads that the less proficient member's score was an equally good predictor of group performance as the better member's. In Comrey's study there was no self or group weighting of the amount contributed by the less-able worker.

It has been found in army-squad research (Havron, 1952) that the squad leader's military knowledge relates positively to ratings of squad performance. The combined ratings of leader and squad performance were also positively related to leader physical ability. The importance of physical ability in army-squad performance was supported by Greer (1955) whose measures were of squad members rather than the squad leader.

On a task having a strong mental rather than physical emphasis (a deceptive horse-trading problem), Thomas and Fink (1961) found the diversity of the members' initial information to be a factor positively influencing quality of group performance. Unfortunately, the relation could not be tested for significance.

Measures of both general and specific ability of individuals have been used to predict small-group task performance. None of the 11 studies reviewed showed a negative relation between member ability and group performance.

Adjustment

Five of the studies in this sample deal with variables relating to adjustment (Cattell, Saunders, & Stice, 1953; Greer, 1955; Havron, 1952; Havron & McGrath, 1961; Haythorn, 1953).

Cattell et al. obtained 15 factors from a 93-variable matrix that included personality measures, observer ratings of member behavior in a group, and group-performance scores. Three of these factors were strongly loaded with adjustment measures. Unfortunately, these factors had few substantial loading-performance variables, although there is a

weak indication that homogeneity in adjustment is conducive to high group productivity. Nothing can be said about the effect of level of adjustment from these factors.

However, Haythorn has used Cattell's Sixteen Personality Factor Questionnaire as a predictor of, among other things, group productivity. He found Cattell's "Emotional Stability vs. General Neuroticism" personality factor to be positively related to group effectiveness, $r = .48$. This personality factor was also related to rating of job completion ($r = .43$). In army rifle squads the combined ratings given to the leader and his squad on performance were significantly related to the leader's adjustment to the Army (Havron, 1952). In another squad study, performance ratings were related to the adjustment of the squad members to the Army (Greer, 1955). Paranoid tendencies of squad members and nervousness were negatively related to these performance ratings. In summing up a series of studies on leader and squad effectiveness, Havron and McGrath (1961) stated that the leader's general army adjustment and low peripheral nervousness and hypertension correlated on the average .25 with squad effectiveness.

Thus, five clear relationships between adjustment measures and group task effectiveness were discernible, and all five of these were positive.

Extraversion

Cattell, Saunders, and Stice (1953) have three personality factors relevant to extraversion in their factor-analytic study of the small group. Again, however, the factors had few performance measures with strong loadings, but these few indicate a positive association between extraversion and high group performance.

The study by Gurnee (1937) cited earlier, in which groups and individuals were compared in ability to learn a maze, is relevant. Using the Bernreuter B_2 scale, he found a positive tendency for groups composed of members who "rarely ask for sympathy or encouragement and tend to ignore the advice of others" to make more errors ($r = .26$). Similarly, using the Bernreuter F_2 scale he found a tendency for groups composed of "nonsocial, solitary and independent" persons to make more errors ($r = .34$). Thus there is some evidence that groups composed of extraverted rather than introverted members perform more effectively.

Dominance

Ghiselli and Lodahl (1958) found that on a cooperative task, groups that contained one subject who tended to be active, reckless, and full of ideas, that is, a dominant person, and two or three average or low dominant subjects were most effective. The more groups approached this "skewed" kind of distribution the better was their performance ($r = .82$). Although the degree of skewness in dominance distribution appears to be related to high group performance, another study has shown the simple variability of the group with respect to individual prominence to be unrelated to group effectiveness (Shaw, 1960). The same cautions mentioned earlier regarding extrapolation from communication-net research are relevant here. (It has, for example, been found by Berkowitz, 1956, that the communication-net design is so constraining that subjects very high and very low in ascendance are indistinguishable when placed in the central position of a wheel-net design.) In summary, two studies dealt with the relation between member dominance and group performance. The results suggest that the way that a group "bunches"

in member dominance may be more important than either the average or the variability of the group's dominance scores.

Authoritarianism

High scorers on this variable tend to seek rules and some form of hierarchical organization in their interpersonal relations. Four studies deal with the relationship between authoritarianism and performance. McCurdy and Eber (1953) constructed groups that were homogeneous in attitude, either all authoritarian or all "democratic." (Authoritarianism was measured by the California F Scale.) They organized some of the groups cooperatively without a leader and some with one of the subjects as absolute leader. Of the groups organized cooperatively, those containing subjects with democratic attitudes made fewer errors in learning patterns of switches to turn on a light than those with authoritarian members ($p < .05$). Apparently a strong group organization acts to suppress differences due to personality, since in the groups with leaders, the members' attitudes did not relate to group effectiveness. Work on other strongly organized groups supports this conclusion. For example, authoritarian attitudes of either the leader or members of groups attempting to write a script for a human-relations film were found to bear little relation to ratings of group competence, productivity, and communication effectiveness (Haythorn, Couch, Haefner, Langham, & Carter, 1956). Similarly, Havron and McGrath (1961) reported that measures of authoritarianism consistently did not relate to army-squad effectiveness.

Finally, Cattell's Personality Factor M, "Bohemianism vs. Practical Concernedness," was found by Haythorn (1953) to relate to rating of group productivity ($r = -.61$) and to rated job completion ($r = -.43$); that is, conservatism of group members was positively related to productivity. In summary, four studies indicate that authoritaranism, as a unitary concept, is unrelated to small group productivity, especially under conditions of high organization where little latitude exists for personality variation to have a great influence on performance procedures.

Other Characteristics

A number of studies deal with the relation between group performance and an aspect of personality that is not categorizable in the previous five groups.

For example, Grace (1954) related the performance of 14 high-school basketball teams with the homogeneity of the members' self-description. The measure was acquired by having each member choose the statement most descriptive of himself and the statement least descriptive of himself from sets of four statements. The degree of conformance (the extent to which members described themselves the way their teammates described themselves) was positively related to number of games won (rank-order correlation = .61).

A series of studies by Fiedler (1960) related group effectiveness to a personality characteristic best described as propensity to discriminate one's work acquaintances. Two of these six studies will be described here. Fiedler had basketball-team members (the same subjects used by Grace) fill out a self-description questionnaire three ways: as the items described them, as they thought their most preferred co-worker would respond, and as they thought their least preferred co-worker would respond. For the group as a whole, members of good teams had significantly greater discrepancy between their ratings of best and worst co-workers

than did members of poor teams. Of the players most chosen as "best co-worker," those on effective teams showed significantly less similarity in their ratings of themselves and their positive choices, and of their positive and negative choices, than did those on less effective teams.

In another study, this time using aircraft crews, the results were in general substantiated, although only for those groups that manifested certain rather specific sociometric characteristics. Although Fiedler was mainly concerned with the influence of the interpersonal warmth or distance of the leader in a group as a determinant of task effectiveness, as he studied more formally organized groups he was forced to control for an increasing number of social determinants, both in the organization in which the group was imbedded and within the group itself.

Another personality trait that does not fit easily into the initial categorization is Schutz's (1958) personal-counterpersonal dimension of his FIRO (Fundamental Interpersonal Relations Orientation) test. An individual is personal to the extent that he will ascribe to such statements as "I like groups where people get personal" and "I like to talk about myself in a group." Using scores on the test, Schutz constructed three types of groups: all were counterpersonal members, all were personal members, or a combination of both counterpersonal and personal members. The counterpersonal groups performed as well as the personal groups on a wide variety of tasks ranging from building a structure to making a group decision about moving chess pieces. However, both the personal and counterpersonal groups performed better than the heterogeneous groups.

Thus, there is a suggestion that three kinds of personality characteristics are related to group performance: The homogeneity of members' self-description, the tendency of the keyman in a group to discriminate sharply his able and less able co-workers or subordinates, and homogeneity among the members in their preferred degree of interpersonal closeness.

The relations of six different personality categories to group performance have been reviewed. The direction of the relationships has usually been clearly indicated, but these relationships are weak for predictive purposes. This weakness in predicting group effectiveness from member characteristics has also been found by McGrath (1962) in his more general review of small-group variables.

To what are we to attribute this general weakness? Social and situational variables interact with personality characteristics and increase "random error" over studies. The influence of one social variable, group organization, has been seen in a number of studies; now an analysis of the task will be used to illustrate the influence of situational demands.

In the Ghiselli and Lodahl experiment (1958) mentioned earlier, the Decision-Making Approach (DMA) scale was administered to student subjects. (This scale had earlier differentiated top executives from middle managers.) The subjects were formed into groups of two, three, or four men to operate two model railroad trains around one track (with sidings) in opposite directions as many times as possible. As noted in the section on Dominance, positive skewness of the group scores on the DMA scale was significantly related to group performance ($r = .82$).

The general design of Ghiselli and Lodahl's experiment was replicated by Porter and Kaufman (1959) and by the author. The two later experiments,

however, used a different task—a fiberboard construction task. In these studies, groups which were positively skewed on the DMA scale were not superior on task performance.

Ghiselli and Lodahl (1958) discovered that groups skewed in DMA scores made few organizational or method changes ($r = -.47$; $r = -.60$); and that groups that made few organizational or method changes performed well on the train task ($r = -.64$; $r = -.55$). It can be inferred that skewness in DMA, by establishing a clear leader who early decided on a course and unwaveringly held to it, resulted in high performance on the task. Ghiselli and Lodahl noted that performance depended less on which of two common solutions was chosen than that one method was chosen and adhered to tenaciously throughout the task.

However, on a task presenting more alternative methods of solution and requiring flexibility, such as building a structure with boards, a tendency to remain with an initial approach would not give the best performance. Thus, a given pattern of scores on an individual characteristic measure was conducive to high group performance on one type of task but unrelated to performance on another type.

The results of these studies and those reviewed earlier in the paper support the continually growing feeling among small-group researchers that some kind of multivariate approach must be used in order to predict either group behavior or behavior in groups. Such a model must be able to account for (*a*) variations in the individuals along relevant personality dimensions at different levels of self-awareness and ego involvement; (*b*) variations in the group's social characteristics such as freedom of interpersonal communication, size, and average attraction toward the group; and (*c*) variations in situational demands on group members, especially task. Task requirements may be predominantly mental, for example, member-usable information, reasoning ability, or immediate memory span; predominantly physical, for example, member strength, fine muscle control, or speed; or predominantly motivational, for example, member resistance to fatigue under conditions of monotony or stress. Moreover, tasks may also vary in the amount of interpersonal cooperation they require of the members.

In conclusion, this review has made clear the fact that within the constraints imposed by situational and social variables, personality characteristics do influence group processes, even those as far removed from the individual member as group task performance.

Specifically, two personality characteristics, ability and adjustment, appear to be fairly consistently related to performance measures. We may conclude that if research on small group performance is to be undertaken, the investigator would be wise to measure both the subjects' general ability and their personality adjustment whether he is interested in their place in a multivariate model or whether he merely wants to be able to control variance due to their influence on some experimental variable.

REFERENCES

Berkowitz, L. Personality and group position. *Sociometry*, 1956, **19**, 210–222.

Calvin, A. D., Hoffman, F. K., & Harden, E. C. The effect of intelligence and social atmosphere on group problem solving behavior. *J. soc. Psychol.*, 1957, **45**, 61–74.

Cattell, R. B., Saunders, D. R., & Stice, G. F. The dimensions of syntality in small groups: I. The neonate group. *Hum. Relat.*, 1953, **6**, 331–356.

Comrey, A. L. Group performance in a manual dexterity task. *J. appl. Psychol.*, 1953, **37**, 207–210.

Fiedler, F. E. The leaders' psychological distance and group effectiveness. In D. Cartwright & A. Zander (Eds.), *Group dynamics*. Evanston, Ill.: Row, Peterson, 1960.

Ghiselli, E. E., & Lodahl, T. M. Patterns of managerial traits and group effectiveness. *J. abnorm. soc. Psychol.*, 1958, **57**, 61–66.

Grace, H. A. Conformance and performance. *J. soc. Psychol.*, 1954, **40**, 333–335.

Greer, F. L. Small group effectiveness. Institute Report No. 6, 1955, Institute for Research on Human Relations, Philadelphia, Pennsylvania.

Greer, F. L., Galanter, E. H., & Nordlie, P. G. Interpersonal knowledge and individual and group effectiveness. *J. abnorm. soc. Psychol.*, 1954, **49**, 411–414.

Gurnee, H. Maze learning in the collective situation. *J. Psychol.*, 1937, **3**, 437–443.

Havron, M. D. The effectiveness of small military units. Personnel Research Section Report No. 980, 1952, Adjutant General's Office, Washington, D. C.

Havron, M. D., & McGrath, J. E. The contribution of the leader to the effectiveness of small military groups. In L. Petrullo & B. M. Bass (Eds.), *Leadership and interpersonal behavior*. New York: Holt, 1961. Ch. 10.

Haythorn, W. The influence of individual members on the characteristics of small groups. *J. abnorm. soc. Psychol.*, 1953, **48**, 276–284.

Haythorn, W., Couch, A. S., Haefner, D., Langham, P., & Carter, L. F. The behavior of authoritarian and equalitarian personalities in groups. *Hum. Relat.*, 1956, **9**, 57–74.

Hoffman, L. R. Homogeneity of member personality and its effect on group problem solving. *J. abnorm. soc. Psychol.*, 1959, **58**, 27–32.

Hoffman, L. R., & Maier, N. R. F. Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *J. abnorm. soc. Psychol.*, 1961, **62**, 401–407.

McCurdy, H. G., & Eber, H. W. Democratic vs. authoritarian: A further investigation of group problem-solving. *J. Pers.*, 1953, **22**, 258–269.

McGrath, J. E. *A summary of small group research studies*. (AFOSR Document No. 2709) Arlington, Va.: Human Sciences Research, 1962.

McKeachie, W. J. Student centered versus instructor centered instruction. *J. educ. Psychol.*, 1954, **45**, 143–150.

Pelz, D. C. Some social factors related to performance in a research organization. *Admin. Sci. Quart.*, 1956, **1**, 310–325.

Porter, L. W., & Kaufman, R. A. Relationships between a top-middle management self-description scale and behavior in a group situation. *J. appl. Psychol.*, 1959, **43**, 345–348.

Rohde, K. J. Theoretical and experimental analysis of leadership ability. *Psychol. Rep.*, 1958, **4**, 243–278.

Schutz, W. C. *FIRO: A three dimensional theory of interpersonal behavior.* New York: Holt, 1958.

Shaw, M. E. A note concerning homogeneity of membership and group problem solving. *J. abnorm. soc. Psychol.*, 1960, **60**, 448–450.

Spector, P., & Suttell, B. J. An experimental comparison of the effectiveness of three patterns of leadership behavior. Technical Report No. AIR-196-57-FR-164, 1957, American Institute for Research, Washington, D. C.

Thomas, E. J., & Fink, C. F. Models of group problem solving. *J. abnorm. soc. Psychol.*, 1961, **63**, 53–63.

(Received December 6, 1963)

Psychological Bulletin
1964, Vol. 62, No. 4, 257–272

DEVELOPMENT OF A DIFFERENTIATED HANDEDNESS[1]

ROBERT D. PALMER

Worcester State Hospital, Massachusetts

This paper is concerned with the intraindividual differentiation in motoric skill implicit in the development of a "dominant" hand. Its aims are threefold. It attempts to establish a developmentally oriented conceptual framework within which to incorporate diverse observations on handedness, to subordinate findings from several sources to this general framework, and to point out some important gaps in the research literature and proffer some suggestions for future research. Expressions of handedness and hand differentiation are considered as complex phenomena, rooted in more general aspects of motor and psychological development, and very probably multiply and complexly determined in their origins. The introduction of a developmental perspective provides a consistent framework for ordering and simplifying certain kinds of data.

Research in the area of hand laterality too often has failed to contribute as fully and consistently as it might to the systematic, cumulative growth of knowledge. This failure can be attributed in large degree to limited theoretical perspectives involving (*a*) an excessive reliance upon empirical and pragmatic distinctions, and (*b*) a compartmentalized view of handedness which fails to recognize that handedness is rooted in more general aspects of motor and psychological development. Because the morphogenesis of handedness is complex and but little understood, it would seem desirable at the present time to attempt to view handedness in as broad a perspective as possible.

[1] This work was supported in part by United States Public Health Service Grant M-5773. The writer prepared this paper while a United States Public Health Postdoctoral Fellow in Clinical Psychology. He wishes to express his indebtedness to Roger Bibace, Alan Clarke, Andrew S. Dibner, Walter Mischel, Leslie Phillips, William Vogel, and, especially, Donald M. Broverman, for many valuable comments and criticisms during the preparation of the paper.

Degrees of Differentiation in Handedness

Handedness as Differentiation

Observation from several sources would suggest that handedness develops from originally undifferentiated movement patterns. Werner (1948) writes that diffuse activity characterizes the earliest stages of human development and distinguishes two types of primitive diffuse movement. One is a global, wholly integrated mass activity. The other is marked by a lack of hierarchic integration or coordination among the different parts of the total movement. Development proceeds, accordingly, in the direction of greater differentiation and greater hierarchic integration. In the process of differentiation, the cephalocaudad and proximo-distad trends of ontogenetic development are considered to be of importance (Carmichael, 1946; Gesell, 1945, 1947; Jackson, 1928; Thompson, 1946). According to the former principle, development of head and trunk regions precedes that of the lower limbs. According to the latter, there occurs, in the process of differentiation, a progressive advance of motor control in

the fetus and infant, from the larger fundamental muscles to the smaller muscles which execute more refined movements. The trunk is innervated before the shoulders, shoulders before arms, arms before hands (Gesell, 1945).

Travis (1931) views expressions of one-sidedness in development as manifestations of a maturational process and as a "differentiation from a whole." Strong, consistent hand preferences do not characterize younger children: Washburn (1929) and others have commented on the relatively symmetrical movements of the extremities in infants. Asymmetric hand movements developing out of the tonic neck reflex (an asymmetric posturing behavior found in the young infant) are described by Gesell and Ames (1947), but these were observed to be unstable and to be readily submerged by symmetrical bilateral patterns. Giesecke (1936), Orton (1937), Lederer (1939), and Hildreth (1949a, 1949b, 1949c, 1950a, 1950b), too, have commented upon the fluctuating, unstable quality of developmentally early manifestations of lateral dominance. Gesell and Ames underscore the need for integration or interrelation of movement as a concomitant to progressively increasing asymmetrical differentiation.

Handedness and Neuroanatomy

This view of handedness or one-sidedness as the product of a developmental process involving gradual differentiation of functions from an initially undifferentiated whole receives some further support from researchers concerned with problems of cerebral dominance and language lateralization. In a recent review of the anatomical literature bearing on human cerebral hemisphere differences, Bonin (1962) observes that reported morphological differences between the hemispheres are minute, are divided about equally between support for dominance of the left and of the right hemispheres, and could not, in his opinion, account for the very great functional differences between the hemispheres.

Penfield and Roberts (1959) find in the genetic literature little convincing evidence for the Mendelian inheritance of handedness, and suggest that man is right-handed out of "pure chance—then: custom and laziness." Zangwill (1960) believes that the hereditary determination of cerebral dominance (for language) is far from absolute, and suggests that at birth the two hemispheres are virtually equipotential in regard to the acquisition of language. Whether this equipotentiality is to be viewed as extending to the mediation of manual functions is left somewhat unclear. Zangwill tends to view both handedness and brainedness as graded characteristics, and would appear to leave the door open for a developmental interpretation of lateralization of function in both areas. Goodglass and Quadfasel (1954) hypothesize two potentially trainable language association systems in the infant, and suggest that after an initial phase of almost equal bilateral participation, a self-reinforcing process causes a slight lead in the language performance of one hemisphere to "snowball into a nearly complete unilateral representation."

Since speech and hand functions are mediated by different areas of the cortex, a finding of equipotentiality for language functions in the infant, while suggestive, would not in itself constitute absolute proof of equipotentiality for motor or for sensory functions. According to Penfield and Roberts (1959), grave injury to speech cortex in one hemisphere in infancy or childhood may cause transfer of speech dominance to the opposite hemisphere, but hand domi-

nance will not be caused to change over unless there is injury to arm areas also.

Handedness as an Index of Development

If the two hemispheres are in fact relatively equipotential at birth with respect to the mediation of manual functions, and one-sidedness is viewed as a "differentiation from a whole," then the degree of lateralization of motor functions observed in the normal individual may be viewed and utilized as an index of development. In this view, an individual with a strongly lateralized pattern of hand skill, in the absence of known brain pathology, must probably be viewed as more differentiated motorically than a more ambilateral or ambidextrous individual, and as having achieved a "developmentally later" level of motor functioning. Let us now consider the possible significance of a differentiated versus undifferentiated dimension of handedness in those areas where traditional interest in handedness has been greatest.

Studies of Handedness

Studies in which handedness has been investigated for its presumed association with mental, cognitive, or emotional defect will be considered under the rubric General Handedness. Those concerned more directly with problems of brain functioning and cerebral dominance will be dealt with under the rubric Handedness and Brainedness.

General Handedness

The overall balance of empirical findings suggests a somewhat greater incidence of left-handedness and/or ambilaterality in certain groups, for example, among stutterers, reading disabilities, mental defectives, and epileptics. References to this sizable literature will be found in Hildreth (1949a, p. 210). Unfortunately, typical studies often fail to separate clearly the influence of left-handedness, on the one hand, and ambilaterality (or ambidexterity or mixed motor preference) on the other. The potential theoretical significance of a differentiated versus undifferentiated dimension of classification of handedness has not usually been seriously considered, and individuals with ambilateral or ambidextrous tendencies have all too often been classified as left-handed. It is therefore difficult to know if the results of these studies are due to "left-ness" per se, or to a relatively less differentiated handedness. Classification is further complicated by the fact that left-handers tend, on the average, to be more variable than right-handers in their hand preferences (Benton, Meyers, & Polder, 1962; Downey, 1933; Humphrey, 1951; Trankell, 1950).

Travis in 1931 showed a clear awareness of the problem, insisting that there were many degrees both of right-handedness and of left-handedness. Hildreth (1949a, 1949b, 1949c, 1950a, 1950b), Subirana (1952), Zangwill (1960), Collins (1961), and others have reached similar conclusions. It has been suggested too, by Blau (1946) and others, that a distinction be made between equal dexterity in both hands (ambidexterity) and an equal lack of dexterity (bilaterality or ambilaterality). While a relative absence of skill in either hand would seem clearly to indicate a less differentiated and developmentally earlier level of motor functioning than a strongly lateralized pattern of hand skill (which implies some proficiency with at least one hand), the developmental status of a high-grade ambidexterity is less clear. A highly developed proficiency with both hands, either for the same tasks (ambidexterity) or for different tasks (a high-grade mixed motor dominance), is probably

fairly uncommon. Certain individuals or occupational groups, however, may be motivated deliberately to train for ambidexterity, or will be selected for their ambidextrous potential (e.g., the dancer or the switch-hitting ball player).

The relationship between handedness and reading problems bears special comment. Dearborn (1933) and Orton (1937) have considered mixed hand dominance to be more of a problem than consistent left-handedness for reading disabilities; data strongly supporting such a position have been reported by Harris (1962). This association is more readily understood in the context of formulations by Tschirgi (1958), Mach (1959), and others suggesting that awareness of spatial position and directionality in space (an ability required for reading mastery) is dependent upon asymmetry of the perceiving system. These authors contend that an animal whose brain is bilaterally symmetrical cannot differentiate between stimuli arriving at homologous points. Werner (1948) and others have documented the egocentric, anthropomorphic characteristics of the spatial world of children and primitives, and suggest that objective space has gradually evolved from this primitive orientation. Kephart (1960) argues that a spatial coordinate system must be established—primarily through motor activity or its observation—within the body itself before directionality in objective space can be appreciated.

In this view, the absence of a consistent laterality should lead to greater directional confusion than a strong right or left skill differential. Benton and Menefee (1957) have demonstrated such a relationship in a group of young children. Benton (1959) postulates a sensorimotor basis for right-left discrimination, in conjunction with certain basic language skills which he believes are necessary for an appreciation of right and left as symbolic concepts. He cites several published cases involving utilization of proprioceptive cues in making right-left discriminations, and points out that children with unilateral motor disabilities (e.g., hemiplegic mental defectives) show relatively good right-left discrimination. Belmont and Birch (1963) report "reliable" right-left discrimination at age 7 and establishment of "consistent" hand preference at age 9. Their conclusion that these two functions are etiologically unrelated must be questioned, for handedness is a matter of degree, and evidence of preferential hand usage can be observed at a very early age. However, the study does point up the need for more definitive research in this area.

Directional confusions lead, in turn, to word and letter reversals, difficulty in reading from left to right, etc. The only difference between a B and a D is one of laterality. Training for consistency of hand usage has sometimes been recommended as a means for establishing lateral directionality and eliminating reading reversals in children (see, e.g., Delacato, 1959). However, if ambilaterality rather than ambidexterity were found to be most centrally involved (indicating a fairly general retardation of motor development), a comprehensive body of techniques for training of basic motor skills, such as those described by Kephart (1960) (possibly though not necessarily including training directly for proficiency of preferred hand usage) might conceivably be more useful than training for consistency of usage.

Considered as a whole, the literature on handedness is susceptible of two major criticisms. The first relates to the too exclusive concern with the pragmatic right versus left distinction, and the failure to grasp the potential utility

of the differentiated versus undifferentiated aspect of handedness as an index of development. One consequence of this restriction has been an overcompartmentalized view of handedness; studies relating development of a differentiated handedness to more general aspects of motor development are needed.

A second criticism concerns the tendency to rely upon questionnaire "consistency" measures of handedness, to the exclusion of behavioral measures of the level of an individual's proficiency with each hand. (With children, the examiner may ask the child to do certain things—throw a ball, hammer a nail—but the subject's score still is based upon the consistency with which he prefers one or the other hand in different activities.) One must ask if consistent preferences necessarily signify a strong unimanual proficiency.

By measuring an individual's proficiency with each hand separately, and in different tasks, more complete information could be obtained than through exclusive reliance on a questionnaire. One could compare lateralized and unlateralized individuals with respect to level of skill with either hand to see how lateralization was in fact achieved, or inquire into the dimensional properties of a domain of hand specialization indices through factor analytic techniques. Appropriate factor scores of hand specialization might well have greater stability and greater predictive value (e.g., for reading problems or language lateralization) than questionnaire or single behavioral measures. Too, it is possible that individuals (or hands) differ in the degree to which they have achieved an asymptote of learning. Those concerned with subjects' "original" handedness might consider possible right-left differences in "limit of learning" and seek to derive ammunition for their theories in a comparison of learning curves for the two hands. The concept of learning limits has received its most explicit use by Ferguson (1954, 1956), who suggested that "abilities," so-called, had reference to overlearned performance at a crude limit of learning.

Consideration might be given to the potential utility of measures of hand differentiation for several other areas of research endeavor. Studies in which the performances of hyperactive and hypoactive children are compared (e.g., Switzer, 1961) or research utilizing the "action-thought" development paradigm (e.g., Phillips & Zigler, 1961) tend (*a*) to equate motor activity with diffuse, affect-dominated or symptomlike behavior, and therefore (*b*) to equate developmental progression with suppression of motor in favor of ideational activity. The progression within the realm of specifically motor development—from affect-dominated to ego-controlled skilled behavior and from diffuse ambilateral to laterally differentiated skilled behavior—is not usually considered. Measures of lateralization in motor functions, as indicators of (motor) developmental level, might conceivably have utility too for sensory-tonic field perceptual research (Wapner, 1962) in which differential behavior on perceptual tasks involving, for example, perception of apparent verticality, is predicted on the basis of subjects' developmental level (particularly in view of the spatial nature of such tasks and the apparent isomorphism of body and external, nonbody, spatial coordinate systems). Too, lateralized motor development undoubtedly has certain implications for the enduring, background distribution of muscular tonus; further research might conceivably demonstrate some equivalence between functional asymmetries of this sort and the experimentally induced, transient tonic

states utilized in sensory-tonic field research.

Handedness and Brainedness

Studies in this second category are concerned with certain established or hypothesized relationships between handedness (right versus left) and various aspects of brain functioning, language in particular. With respect to many motor and sensory functions, activity in the right side of the body would appear to be mediated primarily by the left cerebral hemisphere, and vice versa. For example, if sodium amytal is injected into the carotid artery supplying one or the other side of the brain, a brief contralateral hemiplegia is produced (Wada & Rasmussen, 1960). On this level, then, handedness would seem to betoken a roughly corresponding brainedness. If the hemisphere injected happens to be dominant for speech, a dysphasia is also produced.

For many years it was believed that language and associated functions in the adult were mediated solely or almost solely by the "dominant" or "leading" hemisphere, by which one meant the hemisphere contralateral to the "preferred" hand. In recent years, however, this classic rule of brainedness has been challenged by a number of investigators. Several different positions have been espoused; for example, that both right- and left-handers are left cerebral dominant for language (Roberts, 1951, 1955), or that right-handers are almost invariably left-dominant for language, but that either the right or the left hemisphere (but not both) mediates language functions in the left-handed (Bingley, 1958).

Of especial relevance for the present paper is a position espoused or considered by Conrad (1949), Subirana (1952), Goodglass and Quadfasel (1954), Hécaen and Piercy (1956), and Zangwill (1960)—that left-handers may be less completely lateralized for language, having a greater degree of bilateral representation for speech than right-handers. Since left-handers apparently tend to be somewhat less completely lateralized in their hand preferences as well, and because of the tendency to group ambidextrous or ambilateral individuals with left-handers (e.g., Hécaen & Piercy, 1956), it is conceivable that observed differences in degree of lateralization of language functions are related as much to degree of lateralization of hand functions as to leftness per se. Chesher (1936) presented evidence supporting such a position. Zangwill (1960), working with data supplied by Luria (1947), found that a certain proportion of dextrals (as well as sinistrals) showed incomplete lateralization of the speech mechanisms. Raney (1939) and Rheinberger, Karlin, and Berman (1943), though not dealing specifically with language mediation, demonstrated somewhat greater bilateral asymmetry of electrical brain activity in those individuals with most strongly lateralized hand dominance. A similar study by Glanville and Antonitis (1955) gave negative results, however.

The issues are undoubtedly more complicated than our presentation would imply. Evidence presented by Lansdell (1962) suggests that right-hemisphere dominance for language may constitute a response to early neurological damage to the left hemisphere, and hence that right- and left-hemisphere dominant individuals should not be viewed simply as mirror-image equivalents. Too, an individual's cerebral status almost certainly involves more than a question of degree of lateralization of language functions, or of left- versus right-brainedness. Hécaen and Piercy (1956) suggest that individuals may

vary not only in degree of language lateralization, but also in the diffuseness of language representation within a single hemisphere. They also raise the possibility of different degrees of cerebral focalization for the receptive and expressive components of language mediation. Zangwill (1960) suggests that a greater degree of lateralization of function may occur for more complex or highly evolved aspects of language than for primitive, emotional, or involuntary speech responses. Further, there is in all probability a greater degree of bilateral representation for language or for certain of its aspects in the young child than in the adult, with increasing lateralization or focalization of function occurring in the early years of development. Aphasia follows right-sided lesions more often in children than in adults (Freud & Rie, 1891; Sachs & Hausman, 1926) and is more often transitory. Total removal of the left hemisphere early in life need not preclude the normal development of speech (Carmichael, 1954; Krynauw, 1950). In these respects, and in their greater susceptibility to expressive rather than receptive aphasias (Guttmann, 1942), children may show in their cerebral organization a certain resemblance to left-handed adults (Hécaen & Piercy, 1956).

Relatively little is known of the cortical representation for sensory and motor functions of the hands. However, results of an investigation by Semmes, Weinstein, Ghent, and Teuber (1960) suggest possible parallels between the patterns of hemisphere representation for language and for somatosensory functions. In a predominantly right-handed population the sensory functions of the right hand were found to be more focally represented (in the sensorimotor region of the contralateral hemisphere) than were the corresponding functions of the left hand (which were more widely dispersed within the contralateral, and even, in many cases, the ipsilateral hemisphere). Thus, the pattern of sensory representation for the left hand (among dextrals) appeared to be somewhat analogous to the less lateralized or more diffuse pattern of language representation postulated among left-handers. Since sensory and motor representations in the cortex are grossly similar anatomically, motor as well as sensory functions of the left hand (in dextrals) might conceivably be found to be more diffusely represented.

Semmes et al. (1960) speculate that in the child, lateralization of motor functions occurs first, followed by lateralization of sensory, and of language, functions, respectively. They do not try to account for the genesis of these gradually emerging hemisphere differences, however, or attempt to relate them. Calculations by Bingley (1958) show a clear-cut though far from perfect, statistical association between handedness and brainedness (right versus left). Use of different or more complex models and measures of handedness and brainedness, including, for example, consideration of degree of lateralization of function, might conceivably demonstrate a greater degree of association. Goodglass and Quadfasel (1954), however, specifically reject the notion of a direct causal connection between handedness and brainedness (right versus left). Still, the possibility that handedness in some way influences or determines language dominance, or vice versa, must probably be considered. Such a possibility would seem not unreasonable, for example, if one accepted the notion that primitive language has its roots in gestural com-

munication (Wundt, 1912, pp. 648 ff.; 1916, pp. 58 ff.) or that gestural and verbal symbols subsume equivalent functions in the process of thought (Cushing, 1892), since the preferred hand would ordinarily be the more expressive hand, gesturally. However, purely physiological mechanisms may be responsible for the association.

Handedness and Motor Development

The present section attempts to provide a broader basis for understanding development in handedness, through consideration of more general aspects of the development of manual and motor functions. The failure of handedness research to contribute systematically to a cumulative body of knowledge has been due in some part to its failure to consider the roots of handedness in motor development. Orton (1937), for example, observed that the abnormal clumsiness and motor inaptitude of many ambilateral children involved movements of the body as a whole and not only the complex movements which underlie manual dexterity. Hildreth (1949a, 1949b, 1949c, 1950a, 1950b) has stressed the need for a developmental approach to handedness which would comprehend other manifestations of motor activity. Gesell and Ames (1947), too, view handedness as an integral part of the "total action system" of the organism. Expressions of handedness constitute highly prominent forms of motor behavior, and factors which influence the course of development of motor and manual functions more generally must be considered for their possible impact upon handedness. Understanding of the developing, differentiating aspect of handedness, in particular, enjoins such a consideration.

Drive versus Skilled Aspects of Motor Development

Motility has been considered both for its pleasurable or tension-reducing aspects and for its skilled aspects—its role as an instrument of mastery. References to the expressive or driven quality of primitive motor behavior or to the motor discharge of tension or affect have come from diverse sources. Groos (1898) documented the great force of the play impulse in young animals. The physiologist, Sherrington, suggested that the degree of relief of instinctual tension depended on the degree of motor involvement, with action giving the greatest relief, and thought and fantasy the least (in Freeman, 1948, p. 120). Such a notion is implicit, too, in Freud's (1959a) view of thought as "an experimental way of acting, accompanied by displacement of smaller quantities of cathexis together with less expenditure (discharge) of them." On the basis of controlled observations and experiments with animals and human infants, Levy (1938, 1944) posited the existence of a motor drive. Mittelmann (1954, 1955, 1957) assembled considerable evidence for the existence of such a drive, said to be especially prominent during the second year of life and for several years thereafter due to rapid evolution of postural, locomotor, and manipulative skills.

The pleasurable or tension-reducing ("id") and skilled ("ego") aspects of motor behavior have been differentiated by several observers. Anna Freud and Burlingham (1944) in a vivid description of children under the sway of the motor urge (pp. 14–16), note both the intense pleasure which the child derives from the unrestricted exercise of hands and legs, and the manual and motor skills which develop out of this exercise. Freud (1959a, 1959b) ob-

served that under the supremacy of the "pleasure principle," motility served primarily a discharge function, resulting in affective experience and in an internal alteration of the body without reference to the outer world; with the introduction of the "reality principle," motility was employed in the appropriate alternation of reality, largely under ego control. Similar distinctions were made by Mahler, Luke, and Daltroff (1945). The more skilled forms of motility—being tied, in this view, to the development of the ego, and involving a large learned component—could come to predominate only in ontogenetically later phases of development.

Developmentally early forms of motor and manual behavior appear, then, to be relatively unmodulated and affective in nature and to be geared primarily to problems of tension management created by internal need states. Later forms presumably embody a large learned component, are more responsive to the requirements for action imposed by the external environment, and are more generally under the control of conscious thought processes. Development in hand differentiation would seem of necessity to proceed more or less concurrently with maturation in general motility as here described, since manifestations of handedness constitute prominent forms of motor behavior. Certain implications of this collateral development will be considered below. First, however, let us turn to a consideration of those factors which exert a significant causal influence upon handedness and motor development.

Factors Influencing Handedness and Motor Development

Our concern here is to identify, so far as possible, those genetic and environmental factors which influence the course of development in handedness, and the mechanisms through which they operate. Studies concerned with the origins of handedness fail, for the most part, to deal with the developing, differentiating aspect of handedness, though studies dealing with more general aspects of motor development often deal fairly explicitly with the development of manual functions and skills.

Genetic and Prenatal Factors. The genetic contribution to motor development is but little understood at the present time. According to Anastasi (1958a), hereditary influences may become manifest for the first time at any age, and environmental influences begin to operate long before birth. Sontag (1941, 1950), Pasamanick and Knobloch (1961), and others have emphasized the significance of the fetal environment in determining characteristics of the physiological behavior of the newborn infant. Hence, though infants vary considerably in the amount of motor activity they show at birth (Fries & Woolf, 1953; Irwin, 1932), this variation cannot be attributed solely to innate factors.

In recent years, genetic models of handedness have been advanced by Rife (1950), Trankell (1955), and Merrell (1957). Fuller (1960) observes that handedness (right versus left) has long been a favorite subject for behavioral genetics, providing a convenient testing ground for ideas on nature-nuture interaction. However, identical findings have all too often led to opposite conclusions in the hands of investigators of nativist versus environmentalist orientations, with no critical test available to resolve these differences. That genetic factors are intimately involved in the process of development, including the development of a differentiated handedness, is hardly to be doubted, for most present-day geneticists believe that both hereditary and environmental factors

enter into all behavior. It is probable, however, that the learned and unlearned components of behavior are frequently present as relatively inseparable co-functions, and may therefore be quite difficult to isolate (Anastasi, 1958b; Sperry, 1958). Dennis (1941) and Beach (1955) suggest further that it is impossible at the present time to develop an operational method for distinguishing between learned and unlearned responses, since we have no universal agreement as to what constitutes learning. In view of the many possible patterns of causal sequence by which a gene mutation may effect a change in behavior (Sperry, 1958), the different possible modes or levels of environmental influence on development, and the diversity of ways in which hereditary and environmental factors may interact (Anastasi, 1958b), it would seem important to establish the specific *modus operandi* of genetic and of environmental influences on handedness behavior.

A genetic model to predict right- versus left-handedness (handedness being viewed as a dichotomous variable) clearly should not be expected to account for those differentiated or graded aspects of handedness that have been emphasized in the present paper. A very slight, genetically controlled morphologic imbalance favoring one or the other hemisphere would (along with socialization experiences) be expected to contribute to an individual's favoring one or the other hand, and perhaps ultimately, through more extensive use and learning with the preferred hand, to his becoming predominantly right- or left-handed. This particular model would have to be elaborated somewhat, however, to account for differences in the degree of lateralized function ultimately achieved. Presumably, greater emphasis would now have to be placed upon polygenic inheritance leading to relatively continuous distributions of handedness (Rife, 1940), rather than upon single Mendelian genes. Differences in degree of lateralization might conceivably be attributed to different, genetically given "limits of learning" or "learning efficiency rates" for the two hemispheres, though, as we have seen, little support for such a position can be obtained from present morphologic findings. As suggested in a later section, more indirect mechanisms, learned and/or innate, might conceivably lead to the same phenotypic differences, in spite of relatively equal hemisphere potential at birth. Further investigation of the complex, phenotypic patterning of handedness, including an inquiry into its relationship to more general aspects of motor development, would seem prerequisite to the construction of an adequate developmental (genetic and/or psychogenetic) model.

Environmental Factors. Clear-cut evidence for an environmental influence on handedness, specifically, is difficult to find, though Falek (1959) observed that left-handed fathers produced fewer left-handed progeny than would be expected, apparently because they found sinistrality a disadvantage in their work and trained their children against it. Blau (1946) speculated that many cases of left-handedness were attributable to "emotional negativism."

The influence of environmental factors on the course of development of manual and motor functions more generally, however, has been demonstrated in several studies—most dramatically, perhaps, in the extraordinary motor retardation observed by Spitz (1945, 1947) in his foundling-home children. These emotionally deprived children were found to be seriously retarded in the area of hand skills as well as in locomotor and speech development. Williams and Scott (1953) reported significant relationships between specific child-rearing practices and motor-development scores. Accelerated develop-

ment of manual and locomotor skills under conditions of unrestricted motility and social reinforcement (in a wartime nursery setting) was observed by Anna Freud and Burlingham (1944).

In these latter studies, the environmental influence would appear to have operated primarily by providing, or by failing to provide, opportunities and encouragement for the learning and practice of skilled motor behaviors. There is some evidence, however, for a somewhat different *modus operandi* of environmental influence on motor and manual functions, involving interference with the learning of skilled motor behaviors through the intrusion or threatened intrusion of primitive, drive-linked forms of motility. Motor tics in children (Mahler, Luke, & Daltroff, 1945) and animals (Levy, 1944), elicited by restricted opportunities for motility, constitute one such intrusion. Fenichel (1928, 1945), in particular, has emphasized the role of muscular dystension in the neurotic's line of defense against unacceptable impulses. According to Fenichel, defense processes aim primarily at barring the warded-off impulses from motility, while barring them from consciousness is only a means of achieving this; hence, pathogenic defenses always signify an inhibition of certain movements. Similar notions, less carefully explicated, underlie Reich's (1949) conception of the "muscular armoring" of the neurotic patient; both Fenichel and Reich saw the neurotic as "muscularly dystonic"—the dystonia representing a struggle between motor impulse and a tendency to block the intended movement. According to Reich, hypertonicity or movement inhibitions of the chest musculature (presumably typical in the compulsive character) result in "awkwardness" of the arms and in an inhibition of those arm movements which express "reaching-for-something" or "embracing."

There would appear, then, to be at least two modes through which environmental factors can impede (or facilitate) the course of development of manual and motor functions. On the one hand, the environment may simply fail to provide the individual with sufficient opportunities and encouragement for the learning and performance of (developmentally advanced) motor skills. On the other hand, functional muscular disturbances or motor inhibitions provoked by the threatened eruption into motility of primitive, drive- or affect-related impulses may constitute a direct impediment to the learning and performance of (developmentally advanced) skilled motor behavior. Since primitive, drive-related motor urges are most characteristic of earlier stages of motor development, the existence of functional muscular inhibitions of the sort described by Fenichel and Reich would seem to bespeak a continuing preoccupation with developmentally early forms of motor behavior. The potential impact of drive-related motor behavior on handedness development will be considered below.

Incentives to Specialize and the Problem of Individual Differences

The factors leading to a specifically "unilateral" development in handedness (or brainedness) are but little understood. Since present evidence suggests that the two hemispheres are relatively equipotential at birth, an assumption of (direct) genetic control of the process of lateralization seems unwarranted, and other explanations must be sought.

The idea that lateralized function is somehow more "efficient" is not a unique one. Travis (1931) saw it as providing "required specialized organs," Blau (1946) as making for "economy in development for special capacities," and Gesell (1945) as leading to "effective attentional adjustments." Goodglass and Quadfasel's (1954) notion of

a snowballing "self-reinforcing" process in language lateralization seems particularly apt when applied to the area of handedness: once a young child achieved even a slight preference for one hand in any given (unimanual) task, this dominance should rapidly extend itself to more and more tasks and skills, through transfer of training to common elements of these tasks.

"Efficiency" notions generally do not specify in sufficient detail either the source (endogenous or environmental) or the mechanism of factors pressing for lateralization of functions. Too, they fail to explain why some individuals do not develop a specialized handedness. In the broadest view, of course, the process of differentiation in handedness is probably influenced by any of a number of factors contributing to motor or psychological development or to ontogenetic growth generally. Research in this area is much needed. Several more specific formulations could also be advanced, however, to account for individual differences in degree of lateralized development. Those considered below will have differential utility for psychologists of different theoretical persuasions and interests. They vary in level of generality, and are not to be considered as mutually exclusive.

1. Lateralized development might be seen as a response to environmental demands for skilled performance—for use of the hand as a tool. Efficiency notions themselves seem often to presuppose a commitment to rational, instrumental, environmentally oriented action. Too, this kind of performance will be seen to correspond rather closely with developmentally later, environment-oriented forms of learned, "ego-controlled" motility described by psychoanalytically oriented observers, and seems largely incompatible with developmentally early, more pleasure-oriented forms of motility. An overriding preoccupation with or valuation of the inner world of thought or fantasy, or a strong investment in "intellectual" pursuits to the relative exclusion of action-oriented modes of expression, might in this view preclude attainment of a strongly lateralized handedness.

2. Whereas the above formulation has emphasized the press of the external environment for specialized development, and differential individual responsiveness to this press, the present formulation focuses more upon the internal demand characteristics of primitive drives—the bilateral constraints on motility imposed by instinctual demands for (bilateral) pleasure experience. That is, in the absence of findings to the contrary, one would assume that a movement response on one side of the body would elicit approximately the same amount of undifferentiated pleasure experience or reduction in drive tension as a movement response on the other side; hence, a class of movement responses governed solely by the "pleasure principle" should be relatively bilateral in nature. Excessive continued preoccupation with affective or drive-related forms of motility might in this view preclude attainment of a strongly lateralized handedness.

3. Describing the "intrinsic inhibition" explanation of response decrement, Osgood (1953) observes that the greater the intensity of fatigue at any moment, the greater the tendencies to inhibit the response or to shift to alternate muscle systems (as in carrying one's bag from the station) in order to reduce the total accumulation of psychologically effective fatigue. Eysenck (1955) has postulated the existence of individual differences (presumably innate) in the strength of reactive inhibition and the speed with which it is produced and dissipated. Were an individual more prone than others to develop a strong reactive inhibition following exertion, he might be less able or willing to expend the sustained, con-

tinuous effort necessary to develop automatized skills with one (the preferred) hand. He might instead be more prone either (*a*) to shift hands, and thereby acquire a mediocre level of skill with either hand ("low-grade" ambidexterity), or (*b*) to give up trying altogether, and thereby fail to establish any hand proficiency whatever (ambilaterality in the narrow sense). Were hand specialization found to be a multidimensional phenomenon, it would tend to militate against a fatigue theory of specialization, unless—as suggested, perhaps, by findings of Rechtschaffen (1958) and Becker (1960)—fatigue can itself be conceived as multidimensional in nature.

4. It is possible that some individuals have developed a relatively specialized—or at least consistent—pattern of handedness by virtue of personal rigidities involving an unwillingness to change an established perceptual-motor set or an inability to tolerate or explore the awkward, undeveloped, less-than-best performance of the nondominant hand. Should an individual give temporary ascendance to his nondominant hand (reverse his handedness), he is forced, for the nonce, to take the role of a motorically more awkward or untrained individual. If he is not able, concurrently, to relinquish his usual standards for instrumental achievement, he may find his new identity an unpleasant and conflictual one.

5. Finally, in spite of the relative symmetry of the two hemispheres in gross morphology, one cannot rule out absolutely the possibility that future research will disclose significant native differences in hemisphere structure on a cellular or subcellular level—differences which contribute significantly to manifestations of hand lateralization. Too, while gross morphology cannot explain the general population trend to right-handedness, inherited or congenital morphological differences cannot be ruled out entirely when dealing with the individual case.

Dimensionality in Hand Lateralization

Expressions of hand lateralization are very probably multiply and complexly determined in their origins—a condition which militates against the notion of a unitary process or dimension underlying all manifestations of lateralization. The unlikelihood of finding but a single dimension of specialization in handedness is argued, too, by findings of specialized physical development (Paterson, 1930) or multiple growth factors (McCloy, 1940). The generality of any single measure of lateralization must therefore be seriously questioned. This area is much in need of clarification, being deficient both in theory and in data. A factor-analytic approach could be useful, in view of the few guide lines persently available. However, this would not abrogate the need for a more adequate conceptualization regarding theoretically meaningful dimensions of specialization in handedness. Several possibilities offer themselves in this regard:

1. Measures of hand lateralization involving fairly gross arm-shoulder musculature and movement skills may have somewhat different dimensional properties and significance from measures based upon more discrete hand skills since, according to the proximo-distad principle of development, control of movements of large muscle groups proximal to the body axis tends to occur earlier in ontogenesis than independent movements of more distal parts. It is possible, too, that lateralization of more proximal arm/shoulder movement patterns would create stronger pressures for an accompanying reintegration or coordination of subfunctions than would lateralization of peripheral skills.

2. The degree to which one's measures of hand lateralization draw upon simple, highly practiced or "automatized" abilities (as described, e.g., by Broverman, 1960) has potential theoretical and practical significance, since specialization in handedness might conceivably be viewed as an example of automatization or overlearning of behaviors. In this view the preferred hand may serve as "the instrument" of automatization—the means by which one overlearns many simple, everyday behaviors.

3. Meaningful relationships might well be found to obtain between particular dimensions of hand lateralization and dimensions of ability isolated in factorial studies of motor abilities (e.g., Fleishman, 1958). One will note, however, that the notion of hand differentiation involves reference to a relationship between different subfunctions within the same individual. Quite a different situation prevails in the typical study of motor abilities, wherein unimanual task data are rarely obtained for the nondominant hand, and variables are treated as unitary and undifferentiated.

REFERENCES

ANASTASI, ANNE. *Differential psychology.* (3rd ed.) New York: Macmillan, 1958. (a)

ANASTASI, ANNE. Heredity, environment, and the question "how?" *Psychol. Rev.,* 1958, **65**, 197–208. (b)

BEACH, F. A. The descent of instinct. *Psychol. Rev.,* 1955, **62**, 401–410.

BECKER, W. C. Cortical inhibition and extraversion-introversion. *J. abnorm. soc. Psychol.,* 1960, **61**, 52–66.

BELMONT, LILLIAN, & BIRCH, H. G. Lateral dominance and right-left awareness in normal children. *Child Develpm.,* 1963, **34**, 257–270.

BENTON, A. L. *Right-left discrimination and finger localization.* New York: Hoeber, 1959.

BENTON, A. L., & MENEFEE, F. L. Handedness and right-left discrimination. *Child Develpm.,* 1957, **28**, 237–242.

BENTON, A. L., MEYERS, R., & POLDER, G. J. Some aspects of handedness. *Psychiat. Neurol., Basel,* 1962, **144**, 321–337.

BINGLEY, T. Mental symptoms in temporal lobe epilepsy and temporal lobe gliomas. *Acta psychiat. neurol Scand., Copenhagen,* 1958, 33(Suppl. No. 120).

BLAU, A. *The master hand.* New York: American Orthopsychiatric Association, 1946.

BONIN, G. v. Anatomical asymmetries of the cerebral hemispheres. In V. B. Mountcastle (Ed.), *Interhemispheric relations and cerebral dominance.* Baltimore, Md.: Johns Hopkins Press, 1962.

BROVERMAN, D. M. Dimensions of cognitive style. *J. Pers.,* 1960, **28**, 167–185.

CARMICHAEL, E. A. Hemispherectomy and the localization of function. *Lect. scient. Basis Med.,* 1953–54, **3**, 93–103.

CARMICHAEL, L. The onset and early development of behavior. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1946.

CHESHER, E. G. Some observations concerning the relation of handedness to the language mechanism. *Bull. Neurol. Inst. N. Y.,* 1936, **4**, 556–562.

COLLINS, EMMA. Concept of relative limb dominance. *Hum. Biol.,* 1961, **33**, 293–318.

CONRAD, K. Über aphasische Sprachstörungen bei hirnverlätzten Linkshändern. *Nervenarzt,* 1949, **20**, 148–154.

CUSHING, F. H. Manual concepts: A study of the influence of hand-usage on culture-growth. *Amer. Anthropologist,* 1892, **5**, 289–318.

DEARBORN, W. F. Structural factors which condition special disability in reading. *Proc. Amer. Ass. Ment. Defic.,* 1933, **38**, 266–283.

DELACATO, C. H. *The treatment and prevention of reading problems.* Springfield, Ill.: Charles C Thomas, 1959.

DENNIS, W. Infant development under conditions of restricted practice. *Genet. psychol. Monogr.,* 1941, **23**, 143–189.

DOWNEY, J. E. Laterality of function. *Psychol. Bull.,* 1933, **30**, 109–142.

EYSENCK, H. J. Cortical inhibition, figural aftereffects, and theory of personality. *J. abnorm. soc. Psychol.,* 1955, **51**, 94–106.

FALEK, A. Handedness, a family study. *Amer. J. hum. Genet.,* 1959, **11**, 52–62.

FENICHEL, O. Ueber organlibidinoese Begleiterscheimungen der Triebabwehr. *Int. Z. Psychoanal.,* 1928, **14**, 45–64.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FERGUSON, G. A. On learning and human ability. *Canad. J. Psychol.,* 1954, **8**, 95–112.

FERGUSON, G. A. On transfer and the abilities of man. *Canad. J. Psychol.,* 1956, **10**, 121–131.

FLEISHMAN, E. A. Dimensional analysis of movement reactions. *J. exp. Psychol.*, 1958, **55**, 438–453.

FREEMAN, G. L. *Physiological psychology*. New York: Van Nostrand, 1948.

FREUD, ANNA, & BURLINGHAM, D. *Infants without families*. New York: International Universities Press, 1944.

FREUD, S. Formulations regarding the two principles in mental functioning. In, *Collected papers*. Vol. 4. (Orig. publ. 1911) New York: Basic Books, 1959. (a)

FREUD, S. The unconscious. In, *Collected papers*. Vol. 4. (Orig. publ. 1915) New York: Basic Books, 1959. (b)

FREUD, S., & RIE, O. *Klinische Studie uber die halbeseitige Zerebrallahmungen des Kindes*. Vienna, 1891.

FRIES, M. E., & WOOLF, P. J. Some hypotheses on the role of the congenital activity type in personality development. *Psychoanal. Stud. Child,* 1953, **8**, 48–64.

FULLER, J. L. Behavior genetics. *Annu. Rev. Psychol.*, 1960, **11**, 41–70.

GESELL, A. *The embryology of behavior*. New York: Harper, 1945.

GESELL, A. The ontogenesis of infant behavior. In L. Carmichael (Ed.), *Manual of child psychology*. New York: Wiley, 1947.

GESELL, A., & AMES, LOUISE B. The development of handedness. *J. genet. Psychol.*, 1947, **70**, 155–175.

GIESECKE, M. The genesis of hand preference. *Monogr. Soc. Res. Child Develpm.*, 1936, **1**(5).

GLANVILLE, A. D., & ANTONITIS, J. J. The relationship between occipital alpha activity and laterality. *J. exp. Psychol.*, 1955, **49**, 294–299.

GOODGLASS, H., & QUADFASEL, F. A. Language laterality in left-handed aphasics. *Brain,* 1954, **77**, 521–548.

GROOS, K. *The play of animals*. New York: Appleton, 1898.

GUTTMANN, E. Aphasia in children. *Brain,* 1942, **65**, 205–219.

HARRIS, A. J. *How to increase reading ability*. New York: McKay, 1962.

HÉCAEN, H., & PIERCY, M. Paroxysmal dysphasia and the problem of cerebral dominance. *J. Neurol. Neurosurg. Psychiat.*, 1956, **19**, 194–201.

HILDRETH, G. The development and training of hand dominance. Part I. *J. genet. Psychol.*, 1949, **75**, 197–220. (a)

HILDRETH, G. The development and training of hand dominance. Part II. *J. genet. Psychol.*, 1949, **75**, 221–254. (b)

HILDRETH, G. The development and training of hand dominance. Part III. *J. genet. Psychol.*, 1949, **75**, 255–275. (c)

HILDRETH, G. The development and training of hand dominance. Part IV. *J. genet. Psychol.*, 1950, **76**, 39–100. (a)

HILDRETH, G. The development and training of hand dominance. Part V. *J. genet. Psychol.*, 1950, **76**, 101–144. (b)

HUMPHREY, M. E. Consistency of hand usage. *Brit. J. educ. Psychol.*, 1951, **21**, 214–225.

IRWIN, O. C. Motility in newborn infants. *Proc. Ia. Acad. Sci.*, 1932, **39**, 243.

JACKSON, C. M. Some aspects of form and growth. In W. J. Robbins, S. Brody, A. G. Hogan, C. M. Jackson, & C. W. Green (Eds.), *Growth*. New Haven: Yale Univer. Press, 1928.

KEPHART, N. C. *The slow learner in the classroom*. Columbus, O.: Merrill, 1960.

KRYNAUW, R. A. Infantile hemiplegia treated by removing one cerebral hemisphere. *J. Neurol. Neurosurg. Psychiat.*, 1950, **13**, 243–267.

LANSDELL, H. Laterality of verbal intelligence in the brain. *Science,* 1962, **135**, 922–923.

LEDERER, R. K. An exploratory investigation of handed status in the first two years of life. *U. Ia. Stud. infant Behav.*, 1939.

LEVY, D. M. On instinct-satiation: An experiment on the pecking behavior of chickens. *J. gen. Psychol.*, 1938, **18**, 327–348.

LEVY, D. M. On the problem of movement restraint: Tics, stereotyped movements, hyperactivity. *Amer. J. Orthopsychiat.*, 1944, **14**, 644–671.

LURIA, A. R. *Traumatic aphasia: Its syndromes, psychopathology and treatment*. Moscow: Academy of Medical Sciences, 1947.

MCCLOY, C. H. An analysis for multiple factors of physical growth at different age levels. *Child Develpm.*, 1940, **11**, 249–277.

MACH, E. *The analysis of sensation*. New York: Dover, 1959.

MAHLER, M. S., LUKE, J. A., & DALTROFF, W. Clinical and follow-up study of the tic syndrome in children. *Amer. J. Orthopsychiat.*, 1945, **15**, 631–647.

MERRELL, D. J. Dominance of eye and hand. *Hum. Biol.*, 1957, **29**, 314–328.

MITTELMANN, B. Motility in infants, children, and adults. *Psychoanal. Stud. Child,* 1954, **9**, 142–177.

MITTELMANN, B. Motor patterns and genital behavior: Fetishism. *Psychoanal. Stud. Child,* 1955, **10**, 241–263.

MITTELMANN, B. Motility in the therapy of children and adults. *Psychoanal. Stud. Child.*, 1957, **12**, 284–319.

ORTON, S. T. *Reading, writing and speech problems in children*. London: Chapman & Hall, 1937.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

PASAMANICK, B., & KNOBLOCH, HILDA. Epidemiologic studies on the complications of pregnancy and the birth process. In G. Caplan (Ed.), *Prevention of mental disorders in children.* New York: Basic Books, 1961.

PATERSON, D. G. Personality and physique. In J. A. Harris, C. M. Jackson, D. G. Paterson, & R. E. Scammon (Eds.), *The measurement of man.* Minneapolis: Univer. Minnesota Press, 1930.

PENFIELD, W., & ROBERTS, L. *Speech and brain-mechanisms.* Princeton: Princeton Univer. Press, 1959.

PHILLIPS, L., & ZIGLER, E. Social competence: The action-thought parameter and vicariousness in normal and pathological behaviors. *J. abnorm. soc. Psychol.,* 1961, **63**, 137–146.

RANEY, E. T. Brain potentials and lateral dominance in identical twins. *J. exp. Psychol.,* 1939, **24**, 21–39.

RECHTSCHAFFEN, A. Neural satiation, reactive inhibition, and introversion-extraversion. *J. abnorm. soc. Psychol.,* 1958, **57**, 283–291.

REICH, W. *Character-analysis.* (3rd ed.) New York: Farrar, Straus, & Cudahy, 1949.

RHEINBERGER, M. B., KARLIN, I. W., & BERMAN, A. B. Electroencephalographic and laterality studies of stuttering and nonstuttering children. *Nerv. Child,* 1943, **2**, 117–133.

RIFE, D. C. Handedness with a special reference to twins. *Genetics,* 1940, **25**, 178–186.

RIFE, D. C. An application of gene frequency analysis to the interpretation of data from twins. *Hum. Biol.,* 1950, **22**, 136–145.

ROBERTS, L. Localization of speech in the cerebral cortex. *Trans. Amer. Neurol. Soc.,* 1951, **76**, 43–50.

ROBERTS, L. Handedness and cerebral dominance. *Trans. Amer. Neurol. Soc.,* 1955, **80**, 143–147.

SACHS, B., & HAUSMAN, L. *Nervous and mental disorders from birth through adolescence.* New York: Hoeber, 1926.

SEMMES, J., WEINSTEIN, S., GHENT, L., & TEUBER, H. L. *Somatosensory changes after penetrating brain wounds in man.* Cambridge: Harvard Univer. Press, 1960.

SONTAG, L. W. The significance of fetal environment. *Amer. J. Obstet. Gynecol.,* 1941, **42**, 996–1003.

SONTAG, L. W. The genetics of differences in psychosomatic patterns in childhood. *Amer. J. Orthopsychiat.,* 1950, **20**, 479–489.

SPERRY, R. W. Developmental basis of behavior. In Anne Roe & G. G. Simpson (Eds.), *Behavior and evolution.* New Haven: Yale Univer. Press, 1958.

SPITZ, R. A. Hospitalism: An inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanal. Stud. Child,* 1945, **1**, 53–74.

SPITZ, R. A. Hospitalism: A follow-up report. *Psychoanal. Stud. Child,* 1947, **2**, 113–117.

SUBIRANA, A. La droiterie. *Schweiz. Arch. Neurol. Psychiat.,* 1952, **69**, 321–359.

SWITZER, JANET. Developmental differences in place and name sequence learning in normal, hyperactive, and hypoactive eight- and twelve-year-old boys. Unpublished doctoral dissertation, Clark University, 1961.

THOMPSON, HELEN. Physical growth. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1946.

TRANKELL, A. *Vänsterhanthet hos barn i skolåldern.* (With a summary in English) Helsinki, Finland: Forum, 1950.

TRANKELL, A. Aspects of genetics in psychology. *Amer. J. hum. Genet.,* 1955, **7**, 264–276.

TRAVIS, L. E. *Speech pathology.* New York: Appleton, 1931.

TSCHIRGI, R. Spatial perception and central nervous system symmetry. *Arqu. Neuro-psiquiat.,* 1958, **16**, 364–366.

WADA, J. A., & RASMUSSEN, T. Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance: Experimental and clinical observations. *J. Neurosurg.,* 1960, **17**, 266–282.

WAPNER, S. An organismic-developmental approach to the study of perceptual and other cognitive operations. Paper read at the Martin Sheerer Memorial Meeting, University of Kansas, May 1962.

WASHBURN, R. W. A study of the smiling and laughing of infants in the first year of life. *Genet. Psychol. Monogr.,* 1929, **6**, 396–537.

WERNER, H. *Comparative psychology of mental development.* Chicago: Follett, 1948.

WILLIAMS, JUDITH R., & SCOTT, R. B. Growth and development of Negro infants: IV. Motor development and its relationship to child-rearing practices in two groups of Negro infants. *Child Develpm.,* 1953, **24**, 103–121.

WUNDT, W. *Völkerpsychologie.* Vol. 2. *Die Sprache.* Leipzig, Germany: Englemann, 1912.

WUNDT, W. *Elements of folk psychology.* New York: Macmillan, 1916.

ZANGWILL, O. L. *Cerebral dominance and its relation to psychological function.* London: Oliver & Boyd, 1960.

(Received December 16, 1963)

Psychological Bulletin
1964, Vol. 62, No. 4, 273–286

SEMANTIC SATIATION AND GENERATION:

LEARNING? ADAPTATION? [1]

HARRIETT AMSTER

University of California, Berkeley

This review evaluates the reliability of the phenomena of semantic satiation and the applicability of 3 interpretations. Despite some evidence to the contrary, connotative satiation has been demonstrated in many experiments which involve continuous fast repetition of a word for about 15 sec. Generation, the opposite of satiation, appears more likely to occur when the stimuli are meaningless, when different words are interspersed during repetition, when the repetition period is short, and when the rate of repetition is slow. Associative satiation and its cognitive effects were found to be repeatedly demonstrated. Certain assumptions were made in order that the learning interpretation might account for the 2 types of satiation. The inhibition interpretation was found basically adequate in this respect and compatible with the learning interpretation. A new interpretation of connotative satiation in terms of the theory of adaptation level was proposed. This interpretation could account for a variety of connotative effects including some evidence contrary to connotative satiation and for connotative satiation itself.

The need for a review of semantic satiation was sparked by Yelen and Schulz (1963) who failed to replicate the phenomenon. They also presented a critique of some experimental work on the problem (Lambert & Jakobovits, 1960) and questioned its interpretation in terms of reactive inhibition (Lambert & Jakobovits, 1960). In a rebuttal, Lambert and Jakobovits [2] defended the evidence for semantic satiation and presented additional support for its theoretical interpretation. The present paper is a more extensive review which includes not only semantic satiation, but other effects of repetition on verbal meaning.

Semantic satiation [3] is the loss of a word's meaning which comes about through its continued repetition or prolonged inspection. Introspective reports of satiation were originally studied in Titchener's laboratory (Basette & Warne, 1919; Don & Weld, 1924; Severance & Washburn, 1907) and introspective criteria of satiation have continued to be used by subsequent investigators (Fillenbaum, 1963; Miller, 1963; Smith & Raygor, 1956; Wertheimer, 1958; Wertheimer & Gillis, 1958). However, semantic satiation need not be restricted to phenomenal meaning; it also refers to the loss of specific aspects of verbal meaning, which comes about through repetition.

Two types of semantic satiation will be distinguished: Connotative satiation which involves neutralization of connotative meaning and associative satiation which involves a reduction in the availability of the customary associations to a word. The opposite of satia-

[1] This article was written under the auspices of the Institute of Human Learning which is supported by a grant from the National Science Foundation.

[2] Unpublished manuscript, 1963, entitled "The Case for Semantic Satiation."

[3] The terms satiation and generation are reserved for the effects which result from satiating conditions or procedures, the stimulus arrangements which may predispose the subject to satiation.

tion is *semantic generation,* described by Yelen and Schulz (1963) as an increased connotative extremeness induced by repetition. It implies that words which are slightly positive on a certain scale of meaning will, after repetition, become even more positive, while words which are slightly negative will become more negative. An alternative, *consistent generation,* is that repetition changes connotative polarization in a consistent direction on a particular scale. Accordingly, ratings on a hypothetical scale which has a tendency toward positive generation would become more positive after repetition and similarly, ratings would become more negative on a scale which has a tendency toward negative generation. Consistent generation thus involves shifts in one direction which may result in increasing or decreasing polarity. Needless to say, both generation phenomena are incompatible with satiation.

This paper will review studies of semantic satiation and related phenomena, emphasizing relatively recent studies and those which employ objective indicators of specific aspects of meaning. Individual differences in susceptibility to satiation as reported by Warren (1961), Yelen and Schulz (1963), and Das (1964) will not be discussed as a separate topic. Also, no attempt was made to review the literature on the physiological effects of verbal repetition although such effects are known to exist (Mason, 1941). In the first section of this paper connotative effects of repetition will be described. Evidence of connotative satiation and generation will be presented and various criticisms of existing research will be evaluated in an attempt to establish the empirical status of the phenomena. The evidence for associative satiation and its mediational effects will be presented in the second section of this paper. In the concluding section, two predominant and current theoretical interpretations of connotative satiation will be discussed and an auxiliary interpretation proposed. Some cognitive effects of satiating conditions are incorporated in the final section.

Connotative Satiation and Generation

In the original experiments by Lambert and Jakobovits (1960), words were rated on several scales of connotation representing each of three dimensions of meaning. The dimensions, Evaluation, Activity, and Potency, were those found by Osgood, Suci, and Tannenbaum (1957) to be dominant. Among the scales used to represent the three respective dimensions were the good-bad, passive-active, and strong-weak. Typically, a set of words was rated on six or nine scales. The words were then individually exposed for 15 seconds during which the subject repeated the same word 2–3 times per second. Following this repetition, the word was rated on one scale. The procedure was repeated until all words were rated on all scales. Several different control groups were employed. In one of the experiments (Lambert & Jakobovits, 1960), a peripheral control was used in which subjects repeated GRONY and NUKA before rating NEGRO and CANOE, respectively. Although the motor responses involved in the repetition of GRONY and NEGRO (NUKA and CANOE) are identical, the cognitive responses should have been different. Satiation was obtained in the experimental group which repeated NEGRO and CANOE, but was not obtained in the peripheral control condition, or other control conditions. Thus it appears that the cognitive activity which represents the meaning of the symbol must be con-

sistently called into play during repetition in order for satiation to occur.

In one of their studies, Yelen and Schulz (1963) repeated part of an experiment conducted by Lambert and Jakobovits (1960). Using a different measure, they found in contrast to the earlier investigators, a strong trend ($p < .06$) toward generation. In two other experiments in which control groups were employed, the one a repetition of Lambert and Jakobovits (1960) and the other a partial replication using the same words but different scales, no overall evidence for satiation or generation was obtained. However, some scales were found to show consistent trends toward satiation and others showed consistent generation trends. In a final experiment the procedure was repeated with three generation scales and three satiation scales and the difference between the generation and satiation scales was found to be reliable. From the experiments conducted it was not possible to determine which of the two types of generation might be occurring. Scales related to the same dimension of meaning did not appear to show similar trends toward satiation or generation, but the trend did appear related to the initial level of the ratings of the particular set of words. However, the evidence is entirely inconclusive since initial level and scale have not been independently varied.[4]

Regression Effect

Yelen and Schulz (1963) suggest the possibility that both the satiation and generation tendencies may be accounted for by an artifact, regression, which depends on the fact that, in the experiments reported, the initial ratings of words varied from scale to scale. According to a regression explanation, when initial ratings are near the neutral zone, generation is probable; but when they are near the extremes, satiation is more likely. To control for this effect, appropriate experimental procedure entails selection of words at different distances from the neutral point and an equal number of words above and below the neutral point on a particular scale. They should also be distributed symmetrically about the neutral point and not too close to the extremes of the scale. Consequently, different words would have to represent different scales. The regression criticism may be seen to apply to any study involving words unselected for initial connotation and thus applied not only to the studies by Lambert and Jakobovits (1960) but also to the studies by Yelen and Schulz (1963). Yet, admitting that the regression effect is plausible, the real question is, can it account for the obtained satiation and generation?

In a reanalysis of the data of one of the earlier experiments (Lambert & Jakobovits, 1960), in which stimuli tended to be in the middle of the polarity range, Lambert and Jakobovits[2] found satiation on all scales regardless of initial level. Connotative satiation has also been obtained by others (e.g., Das, 1964; Kanungo & Lambert, 1963a, 1963b). Messer, Jakobovits, Kanungo, and Lambert (1964), using different words and scales, but still not controlling for initial levels on the scales, also reported that satiation was obtained on all scales. However, based on their reported means, some evidence of a regression effect can be observed. The rank-order correlation between the

[4] An attempt has been made to vary initial level and scale independently but no significant shifts in connotation were observed despite the fact that the basic satiating procedure (Lambert & Jakobovits, 1960) was replicated (Robert Olton, personal communication, 1964).

magnitude of the satiation effect and the initial ratings is found to be $+.42$ ($p < .05$). The most salient feature of these experiments is that control groups were employed and without invoking concepts such as satiation, it is impossible to explain how a reliable regression effect could consistently occur under the experimental satiation treatments and not under control treatments. Thus the regression hypothesis is quite likely to account for some of the variance in the findings, but does not cast doubt on the existence of connotative satiation.

Differences in Method

In order to understand the satiation phenomenon we must discover why in one case directly opposite results were obtained by two investigators in which the same materials and similar procedures were employed (Lambert & Jakobovits, 1960; Yelen & Schulz, 1963). Although the possibility exists that the subjects represented different populations, there is a strong probability that procedural discrepancies account for the differences, or that the differences may be attributed to the use of different measures of connotative change (Lambert & Jakobovits [2]). Lambert and Jakobovits (1960) used a polarity difference measure while Yelen and Schulz (1963) used a mean difference score. To illustrate the differences between the two scores consider the example supplied by Lambert and Jakobovits.[2] With three initial ratings of 2, 7, and 3 and final ratings of 4, 5, and 5 one obtains a mean difference score of $\frac{4+5+5}{3}$ minus $\frac{2+7+3}{3} = +.67$. This would be called generation since the mean of the scores was above neutral at the outset. If the mean had been below neutral at the outset, it would have been called satiation. Using a polarity difference score, the scores are first transformed into polarity scores, the deviation between the absolute value of the score and the neutral point. One obtains a difference $\frac{0+1+1}{3}$ minus $\frac{2+3+1}{3} = -1.33$ which indicates satiation. Thus whether satiation or generation is observed for the mean difference score depends on the direction of the shift on a particular scale, while the polarity score is independent of this factor. Thus the measure of both satiation and generation as previously defined is the polarity score while the mean difference score more clearly measures consistent generation. The latter can only be used to measure satiation and generation if words are preselected with respect to symmetry of initial values about the neutral point or the point under consideration on each scale.

Although it is theoretically possible that the discrepant results may be accounted for by the different measures employed, it is of interest to obtain some empirical evidence. Consequently, data obtained by Hodge and Battig [5] employing the same words and scales as Lambert and Jakobovits (1960) and Yelen and Schulz (1963) were analyzed by both measures.[6] The experiment (Hodge & Battig [5]) will be described before the results of this analysis are presented.

On the hypothesis that the contrasting results obtained by the two sets of investigators (Lambert & Jakobovits, 1960; Yelen & Schulz, 1963) may be

[5] Unpublished manuscript, 1963, entitled "Effect of Rate of Repetition upon Verbal Satiation."

[6] Scoring by the two measures was suggested by William Battig who very kindly provided the data (see Footnote 5) for this purpose.

attributed to differences in the rates with which the words were repeated, Hodge and Battig [5] compared the effects of fast and slow repetition on satiation-generation. The subjects were 45 undergraduate men from the University of Virginia. Under fast repetition subjects were instructed to repeat the word as fast as they could (Yelen & Schulz, 1963) and under slow repetition they repeated the word 1.5 times per second which is considerably slower than the 2–3 per second rate employed by Lambert and Jakobovits (1960). For the control group which repeated unrelated words, approximately half of the subjects were given the fast rate and the other half were given the slow rate. The three satiation scales differed significantly from the three generation scales in the change in ratings from pre- to postrepetition. The F ratio for scale-type was 4.10 for the polarity score and 17.14 for the mean difference score ($df = 1/42$, $p < .05$ and $< .01$, respectively). However, the differences as a function of rate of repetition were not statistically significant. The F ratio was 2.29 for the polarity score and 1.62 for the mean difference score ($df = 2/42$, $p < .11$ and *ns*, respectively). The means for both measures showed a strong trend toward satiation after fast repetition and a very slight trend toward generation after slow repetition. For both measures, only the satiation scales showed the trend toward satiation under fast repetition and only the generation scales showed the trend toward generation under slow repetition. For the polarity score there was no difference between the two types of scale under the control condition, but for the mean difference score the difference between the two types of scales was almost twice as great under the control condition as under any of the other conditions. Consequently, the differences between the satiation and generation scales may be due largely to regression effects and are not related to the effects of the experimental condition. Further, no evidence of generation and some evidence for satiation were obtained under the fast repetition (satiation) condition. Despite lack of statistical significance, these results seem inconsistent with those of Yelen and Schulz (1963) and consistent with those of Lambert and Jakobovits (1960). The two measures gave slightly different results in the predicted direction; with the polarity score the satiation effects were larger and with the mean difference score the generation effects were larger. However, since the generation trend only occurred under slow repetition, no doubt was cast on the validity of satiation. But Das (1964) found satiation to be quite small despite a 40-second period for repetition, and Yelen and Schulz (1963) did not find a difference between massed and distributed repetition as expected.

Consistent Generation

Despite its merits as a measure of satiation and generation, the polarity difference score may mask consistent generation. Direct evidence for consistent generation stems from the work of Johnson, Thomson, and Frincke (1960) who presented meaningless and connotatively neutral verbal materials to be rated for goodness before and after repetition. Repetition involved explicit pronunciation of different meaningless words presented for a varying number of trials in a scrambled sequence. The hypothesis of generation formulated above would not imply a consistent increase or decrease in connotation as does consistent generation. In accordance with the latter, items which were equal and neutral in rated goodness were observed to increase with fre-

quency of repetition in a statistically reliable manner (Johnson et al., 1960). However, the assignment of words to frequency conditions was not counterbalanced in this study nor was a control group employed. Thus the conclusion may have been an artifact of the particular assignment of words to conditions. Consequently, as part of a larger study, this experiment by Johnson et al. (1960) was repeated by Amster and Glasman (1965) with the addition of a control group and other modifications in procedure. The results were replicated and, in addition, no effect of repetition was found on the variability of the ratings. If generation had occurred, some of the relatively neutral words would have become positive and others negative, producing an increase in the variability of the set of words. Although no trends were found for the strong-weak scale, consistent generation occurred especially for neutral meaningless material on the good-bad scale, and it tended to characterize the active-passive scale. These scales were two of the scales which Yelen and Schulz (1963) singled out as generation scales although they preferred to attribute generation to a regression effect. In the case of the studies by Johnson et al. (1960) and Amster and Glasman (1965), regression could not explain the results because all stimuli were at the outset, neutral in goodness, although they were mildly positive in activity.

Connotative Satiation and Generation Reconciled

The fact that consistent generation has been observed does not preclude the possibility that satiation may exist also. For one thing, there were procedural differences. The major differences between the satiating conditions and the procedure employed by Johnson et al. (1960) and Amster and Glasman (1965) were (*a*) higher frequencies of exposure were employed in the case of the satiating conditions; (*b*) for the satiating conditions the same word was repeated continuously while in the latter case the 20 stimuli were interspersed during repetition; (*c*) for experiments on satiation the connotations of the words were different at the outset, but in this case they were highly similar; and (*d*) for satiation, words were employed while both words and syllables were used in two of the other experiments (Amster & Glasman, 1965). Satiation was observed under satiating conditions for words while consistent generation was observed under the other conditions mainly for meaningless materials. In an experiment by Jakobovits and Lambert (1964) generation or consistent generation was found for objects while connotative satiation was found for their corresponding names.

It would seem reasonable to view the effects of frequency on connotation as falling on a continuum. For example, one possibility is that for meaningless materials, consistent generation increases with frequency up to a point, but it tapers off and reverses to satiation beyond that point. However, the placing of generation and satiation on a continuum awaits the investigation of the two phenomena under parallel conditions and through a sufficient range of frequency to allow the demonstration of both. In future research it would be desirable to use initial stimulus value as a variable, to use different scales for each dimension, and to use a wide variety of different words. If control words are employed, their characteristics relative to the other words such as length, meaningfulness, pronunciability, and rated level of connotation should be taken into account.

In sum, connotative satiation has been repeatedly demonstrated (e.g., Kanungo, Lamber, & Mauer, 1962; Lambert &

Jakobovits, 1960; Messer et al., 1964). Doubtful support has also been reported (Floyd, 1962; Hodge & Battig[5]) while results inconsistent with the postulation of satiation for all scales have been reported in only one study (Yelen & Schulz, 1963). Their demonstration of consistent differences among scales with respect to the tendency for satiation to occur has been found reliable (Hodge & Battig[5]), but their finding that generation occurs in generation scales when conditions should promote satiation, was not replicated (Hodge & Battig[5]). However, on the basis of the one replicated study (Yelen & Schulz, 1963) the possibility remains that generation or consistent generation does occur for some scales under satiating conditions, and it is quite possible that whether a scale yields satiation or generation, depends on the initial values of the words on the scales. The procedures of Yelen and Schulz (1963) and Hodge and Battig[5] do not permit separation of generation and consistent generation. Generation has been observed among one type of bilinguals in a somewhat different situation from those reported in this section (Jakobovits & Lambert, 1961), while consistent generation has been observed as a more general phenomenon. Consistent generation occurs primarily in meaningless items, on certain scales, and when assorted items are interspersed during repetition (Amster & Glasman, 1965; Johnson et al., 1960). It may also occur for words when the rate of repetition is low (Hodge & Battig[5]).

In order to draw definitive conclusions concerning the relationship between satiation and generation, new investigations should sample stimulus values on each dimension which is studied. As mentioned earlier, these values should be symmetrical about the neutral point or about points which are being studied. The most practical procedure involves using different words to represent each scale since it is almost impossible to control the values of a set of words on several different scales at the same time. In addition, it is particularly critical that the particular selection of stimulus words be varied in view of the fact that the same words have been repeated in all too many experiments. Characteristics of the words (e.g., part of speech, word length, concreteness, etc.) should be controlled or studied since they may affect the susceptibility of the words to semantic satiation (Wertheimer & Gillis, 1958), and the sequence in which the words are presented should be completely counterbalanced. Both measures of connotative change, polarity and mean difference, must be employed in order to determine the existence of the two types of generation, unless the words are experimentally preselected on some basis which will permit determination of satiation and generation from the mean difference measure. In addition, results from the various scales should be presented separately since scales may vary in the rate with which changes occur and in their relevance to the meaning of the particular words employed.

Associative Satiation

Several studies indicate that continued elicitation of a word leads to changes in the associations given to that word. Employing a subjective criterion of satiation in which the subject terminated the satiation period by his report that loss of meaning had occurred, Smith and Raygor (1956) found that following prolonged visual exposure to a word, word-association responses were more rare than those following brief exposure. A similar effect has also been demonstrated by a cross-satiation procedure (Fillenbaum, 1963), which involves measurement of satiation following repetition of a synonym. The

main result (Fillenbaum, 1963) consistent through five experiments, was that synonym-satiated words evoked associations which were less frequent than either associations to self-satiated words or control words which were satiated on irrelevant words. With one exception the differences between self-satiated words and control words were considerably smaller and less reliable than the differences between either of these and the synonym-satiated words. The exception involved repetition of a word for a very brief 4 seconds. In this case, self-satiated words yielded associations which were significantly more frequent than associations to control words. In two experiments, one, like Smith and Raygor (1956), employing a subjective termination of the period of satiation, and the other employing a criterion of a fixed number of responses, there was a clear trend for the self-satiated items to evoke associations which were less common than the control items. In two experiments involving repetition of a word for 1 minute and 3 minutes, respectively, the differences between self-satiated words and control words were small and not significant. In conclusion it appears that both synonym-satiation and self-satiation clearly result in a reduction in the dominance characteristic of associative responses but the effects of synonym-satiation are larger and more readily demonstrable.

Further support for associative satiation stems from the work of Kanungo and Lambert (1963a, 1963b) who found that reliably fewer relevant associations were evoked after continuous repetition. Since the number of irrelevant associations was increased in the satiation condition, the total number of associations was not simultaneously reduced by continuous repetition. Indirect support for associative satiation derives from a study (Jakobovits, 1962) of the effect of the satiating treatment on verbal concepts which are required for forming a concept in a subsequent task. Compared with control subjects, those to whom the satiating treatment was administered made more atypical classifications in the later task.

To evaluate the theoretical proposition that semantic satiation is a cognitive form of reactive inhibition many cognitive effects of satiation have been studied. Some of these studies which have been described in a more detailed review of this topic (Lambert & Jakobovits [2]), will be briefly mentioned in this section. No attempt is made to evaluate their methodology.

The effectiveness of a potential mediator was hypothesized (Jakobovits & Lambert, 1962a) and found to be reduced by prior satiation. As materials for a paired-associate learning task they used word pairs for which there was a potential verbal mediator. For example, soldier is a potential mediator for the word pair ARMY-SAILOR. It was found that paired-associate learning was hindered by satiation of the potential verbal mediators. The satiating treatment has also been shown to produce a decrement in paired-associate learning (Kanungo, Lambert, & Mauer, 1962). In addition, for highly meaningful words, the decrement in learning occurs when either the S or the R terms are satiated, but for low meaningful words the decrement only occurs when the S term is satiated (Kanungo & Lambert, 1963a, 1963b). For recall, a decrement occurs only after satiation of the S term (Kanungo, Lambert, & Mauer, 1962). It has also been found, that satiation of one of the numbers to be added, lengthened the time required to add the numbers (Jakobovits & Lambert, 1962b). In this case, satiation of part of a stimulus interfered with the production of a habitual response.

The notion that mediational processes are inhibited through satiation was

studied in bilinguals by Jakobovits and Lambert (1961). Among the bilinguals who were assumed to have a unitary mediational process for the two languages, repetition of words in one language induced satiation in the equivalent word in the alternate language. However, among bilinguals who were assumed to have separate mediational processes, generation seemed to be induced by the same procedure of cross-satiation. In another interesting study of mediation (Miller, 1963), activity relevant to the meaning of the repeated word, for example, PUSH, PULL, was introduced. It was hypothesized that accompanying repetition with activity consonant with the meaning of the word would stimulate the mediational process. Accordingly, such activity has been found to retard satiation (Miller, 1963).

Interpretations of Connotative Satiation and Generation

There are three interpretations which will be described, and all of them could be valid. The distinction between connotative satiation and associative satiation is particularly germain to this discussion since it is possible that the two types of satiation may, under certain conditions, be quite independent. The effects of satiation may relate to any or all aspects of verbal meaning, but the two that have been emphasized are connotative meaning on the one hand, and associative and mediational aspects of meaning on the others.

Learning Interpretation

According to the learning or counter-conditioning interpretation (Kanungo & Lambert, 1963a, 1963b), a tendency is learned for the repeated word to become associated with itself and thus to interfere with established stimulus-response connections. Accordingly, one might expect that the number of associations elicited by a particular repeated word might begin to decline after a few repetitions, but only relevant associations do in fact decrease (Kanungo & Lambert, 1963a, 1963b). However, as we have seen, there is considerable evidence to support the theory that the common associates are inhibited through satiation (Fillenbaum, 1963; Kanungo & Lambert, 1963a, 1963b; Smith & Raygor, 1956). In addition to evidence for associative satiation, generation is expected in that the associations to a word would become increasingly available when the word is first presented and before it becomes substituted for other associations. Evidence for associative generation after a brief exposure supports this (Fillenbaum, 1963). Moreover, most of the evidence for cognitive effects of satiation reviewed in the preceding section strongly support the learning interpretation.

To extend the learning interpretation to account for connotative satiation, it is reasonable to assume that the connotation of a word depends on the associations it evokes. It is known that there is a high correlation between the connotation of a word and the connotation of its associates (Staats & Staats, 1959b). The dependence of connotation on associations may be related to the fact that the connotational scale terms are associates of the particular word (Bousfield, 1961), or there may be generalization of connotative responses from word associates to the word itself. In any case, if richness of connotation is dependent on the number of associations produced, one would expect connotation to be richer at the outset and to decline with frequency of repetition or exposure. This is attested to by the possible evidence for generation with short exposures of the stimulus (Hodge & Battig [5]) and the evidence for connotative satiation (e.g., Lambert & Jakobovits, 1960). In ad-

dition, if the connotation of a word depends on the connotations of particular word associates, words which have predominantly positive associates would be judged increasingly more positive with increased frequency of usage in contexts in which associates continue to be elicited. Since more words are positive than negative in evaluative connotation (Johnson et al., 1960), associates should, on the average, be positive. Thus evaluative connotation should and does increase up to a point with repetition when the situation does not promote the substitution of the word for its word associates (Amster & Glasman, 1965; Johnson et al., 1960). It may be noted that in these experiments the procedure differed from the satiating procedure in that words were interspersed during repetition.

Inhibition Interpretation

Based on the work of Hull (1943) and Osgood (1953), satiation has been interpreted (Lambert & Jakobovits [2]) as a cognitive type of reactive inhibition which is generated by the continuously repeated elicitation of the representational mediation responses assumed to be associated with the meaning of the repeated word. This reactive inhibition temporarily reduces the availability of these representational meaning responses (Lambert & Jakobovits, 1960). This interpretation was modified by Jakobovits (1962) who considered satiation a special case of extinction. According to both of these interpretations, repetition of a word inhibits it and also the associative and connotative responses. This explanation thus accounts for the known facts of connotative satiation, associative satiation, and its cognitive effects. In addition, it accounts for the generation facts if it is assumed that early in learning habit strength increases at a more rapid rate than reactive inhibition. An implication of the inhibition interpretation is that connotative satiation should generalize to semantically similar and associated words. Although no support for this implication was obtained by Floyd (1962) for connotative satiation, some of the results by Fillenbaum (1963) involving associative satiation are consistent with it; namely, that satiation of synonyms was observed.

The main difference between the learning and inhibition explanations is that in the former case satiation involves the facilitation of the response which matches the verbal stimulus and substitution of the word itself for the habitual responses to that stimulus, and in the latter case, it involves the inhibition of the word. Only the latter explanation is compatible with the reduced availability of a satiated word during paired-associate learning (Kanungo, Lambert, & Mauer, 1962). Both learning and inhibition could occur as part of the same explanation. For both explanations, connotative satiation and associative satiation are results of the same inferred events and are thus interrelated.

Adaptation-Level Interpretation

A third interpretation of connotational changes which occur as a function of frequency is based on the adaptation-level formulation (Helson, 1947, 1964) and is being applied mainly to connotative satiation and generation. This formulation provides that judgments are made with respect to internal norms, adaptation levels, which can be objectively and quantitatively specified. The value of these adaptation levels depends primarily on the values of all the stimuli in the situation and the stimulus being judged. Adaptation-level theory has been applied to a wide variety of stimuli (Helson, 1964) and was proposed to account for judgments of connotation in verbal stimuli by

Podell (1961). According to this formulation, when a stimulus in the absence of any context other than itself is repeated continuously, the adaptation level shifts gradually during repetition or exposure until it is equal to the value of the stimulus. Thus, objects viewed in the absence of context would gradually appear neutral. For example, red objects viewed continuously in red illumination lose their redness and appear neutral. Similarly, a stimulus should tend increasingly to be judged as neutral on all dimensions of connotation as frequency of exposure in a uniform context is increased. An implication of the theory is that repetition of different words having the same connotation on a particular scale should have the same effect on that scale as repeating the same word because the connotative contexts would be identical in both cases. Thus adaptation-level theory may adequately explain connotative satiation and this implication points up the need to obtain appropriate control words for satiation experiments.[7] Ideally, the irrelevant control word should be equal to the satiated word on all connotative scales. Since this condition can never be met except in the case of synonyms, it would be necessary to measure satiation separately on each scale and employ a different set of satiated and control words for each scale. Curiously enough, connotative satiation should occur for both the control and the experimental conditions with respect to the scale for which the two types of words were matched. However, connotative satiation should obtain for the experimental words relative to the other scales and associative satiation should obtain only for the experimental words, according to this view.

There is no doubt but that the connotation of the context is an important determinant of the connotation of particular verbal items. According to adaptation-level theory the connotation of a word would tend to shift toward the connotations of the contextual words if it is on the same side of the neutral point as that set of words (Helson, 1964). This phenomenon of assimilation would account for generation under certain conditions and for connotative changes during the acquisition of connotative meaning. It has been clearly established that in addition to other aspects of meaning, words acquire the connotative meaning of words with which they have been systematically presented (Staats & Staats, 1959a, 1959b). The adaptation-level explanation may be considered as an auxiliary to those previously suggested (Staats & Staats, 1959a, 1959b). In contrast to assimilation, if the connotation of a word and a set of contextual words are on opposite sides of the neutral point, contrast should occur. Amster (1964) found evidence for verbal contrast under such conditions. In this experiment the connotation of the context affected both evaluative judgments and recall, but the two types of effects were mainly independent. The adaptation-level formulation permits the possibility of relative independence of connotative and associative changes. Although both types of meaning-change appear to occur under the same satiation conditions, it is possible that even here there may be independence. For example, Kanungo and Lambert (1963b) found no change

[7] For example, repetition of GRONY which we would assume to have neutral connotation should provide an adaptation level which is approximately neutral. The rating of NEGRO on this basis should lead to a connotatively polarized rating. NEGRO should appear polarized in contrast with GRONY which exemplifies the adaptation level. However, when NEGRO is repeated, the adaptation level shifts toward the connotation of NEGRO and away from the neutral value of the scale. Thus NEGRO should appear less polarized relative to the adaptation level which would approach the value of NEGRO.

in connotative meaning for low meaningful items, but they did find cognitive effects of satiation for them.

The fact that generation seems more likely to occur than satiation at low levels of practice may be explained by application of adaptation-level theory. We have assumed that adaptation takes place with repetition such that in time the value of the adaptation level approaches the value of the repeated stimulus. Satiation is thus a special case of adaptation. However, we will now ignore the effects of continued repetition and consider the factors which affect the judgments after one repetition. Assuming that the original stimulus ratings were made with reference to an adaptation level which coincided with the conventional neutral point (AL_1), one can make approximations of the ratings of the stimulus from the knowledge of the mean of the stimuli in the set (AL_2) and the value of the particular stimulus (S). For example, if AL_2 is above AL_1 and the stimulus is below AL_1 the rating of the stimulus should show an assimilation effect which in this case would be interpreted as generation. Specifically, let $AL_1 = 2$, $AL_2 = 3$, and $S = 1$ on an abstract scale. Accordingly, the original rating of S was 1 unit less than AL_1. After repetition, the AL becomes AL_2, and S is rated 2 units less than AL. The S has thus become more distant from the neutral point. If, however, the stimulus was below AL_1, and AL_2 was between AL_1 and the stimulus, a contrast effect should occur which would be called satiation. Let $AL_1 = 3$, $AL_2 = 2$, and $S = 1$. The original rating of S was 2 units below neutral, but the second rating was only 1 below neutral. Also, whether satiation or generation will occur would depend strongly on the magnitude of the original rating of S. It is more likely to be interpreted as satiation when the original rating is extreme. Consider a specific case: Let $AL_1 = 4$, $S = 2$, and $AL_2 = 1$. The original rating would have shifted from 2 below to 1 above neutral. In this case, the contrast effect would be called satiation. Theoretically, this state of affairs should occur at the outset of repetition and should decrease with repetition or repeated exposure. These considerations should pertain to both control and experimental conditions. Ideally, differences between control and experimental groups would be attributed to the repetition of the critical stimulus in the experimental group. However, unless the control words are perfectly matched with the experimental words, at the second presentation of a particular stimulus, the two groups would differ with respect to the values of the stimuli which have been in the prior context and determine the adaptation level. Thus, the fact that generation appears to occur at low levels of practice, at slow rates of repetition, and for certain scales, even under control conditions, is consistent with this formulation.

That consistent generation occurs mainly for meaningless but not meaningful words (Amster & Glasman, 1965) is consistent with the adaptation-level formulation. Since the total frequency of the stimulus in the history of the subject would be so much smaller in the case of meaningless materials, they would be expected to undergo larger changes in connotation with relatively small amounts of repetition. The occurrence of consistent generation may be explained with the aid of additional hypotheses. For example, it is believed that increased familiarity is itself a source of positive value (Postman, 1953) and this in turn may even generalize to other scales. Similarly, increasing pronunciability which occurs with repetition may also be a source of value or pleasantness and might arouse other connotations. All of these factors

could contribute to the observed generation in meaningful materials and an adaptation-level prediction would be that continued repetition of the stimuli would lead to satiation on all scales including those on which the generation had occurred.

Adaptation-level theory suggests a number of hypotheses which might explain the fact that satiation has been found to occur for some scales of meaning and not for others. The first is that when stimuli have extreme values on certain scales, more time is required for adaptation to occur than when the values are less extreme. However, the nature of the function might be such that greater amounts of change would occur at the outset and the rate would taper off with increased repetition. Consequently, a shift toward satiation which is a particular instance of adaptation should be more readily observed in stimuli having relatively extreme values at the outset. Second, the rate of adaptation to a word on a scale might depend on the relevance of that scale to the meaning of that word. If the responses evoked during repetition do not readily evoke responses which will generalize to the scale terms, it is likely that adaptation may be slow or nonexistent. The third is consistent with the possibility discussed above that the adaptation level on a particular scale shifts in position between the time of the initial rating of the word and the time it is repeated. If the shift in rating brought about by this effect was large, it is possible that it would be dominant over the shifts in adaptation which occur over time. In other words, the resulting judgment would reflect a balance between the two factors at a given point in time and make it relatively difficult to observe satiation when the shift was large. "Pure" satiation would only be evident after extremely long exposure to a stimulus. The final hypothesis is somewhat speculative, but deserves serious consideration. It concerns the possibility that there is interaction between the values of a stimulus on various scales, such that the fact that a word has an extreme value on one scale might effect the rate of adaptation on other scales for which the values might be less extreme. Similarly, the values of other words on one scale might affect the value of the particular word on a given scale. This implies that the various adaptation levels would interact.

Associative satiation is also compatible with adaptation-level theory, and one may conveniently invoke the cognitive inhibition interpretation in order to develop a complete explanation of the phenomena of satiation. However, when adaptation level is employed as part of the explanation, connotative shifts are not a necessary condition for associative changes. If the control words are properly matched with the experimental words, connotative satiation would be observed for both groups but associative satiation would only be observed for the experimental group. The correspondence between the connotative and associative changes is yet to be observed in situations involving a variety of verbal contexts. Undoubtedly, the meaning of the context, the contiguity between the word and the contextual words, and the relative numbers of contextual and critical words, are all crucial variables in determining the effects of repetition on the meaning of a verbal item, and much experimental work on these variables remains to be done. Moreover, the relationship between connotative and associative effects may be more complex than has previously been assumed.

REFERENCES

AMSTER, HARRIETT. Evaluative judgment and recall in incidental learning. *J. verbal Learn. verbal Behav.*, 1964, in press.

Amster, Harriett, & Glasman, L. D. Verbal repetition and connotative change. *J. exp. Psychol.*, 1965, in press.

Basette, M. F., & Warne, C. J. On the lapse of verbal meaning with repetition. *Amer. J. Psychol.*, 1919, **30**, 415–418.

Bousfield, W. A. The problem of meaning in verbal learning. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961.

Das, J. P. Hypnosis, verbal satiation, vigilance, and personality factors: A correlational study. *J. abnorm. soc. Psychol.*, 1964, **68**, 72–78.

Don, V. J., & Weld, H. P. Lapse of meaning with visual fixation. *Amer. J. Psychol.*, 1924, **35**, 446–450.

Fillenbaum, S. Verbal satiation and changes in meaning of related items. *J. verbal Learn. verbal Behav.*, 1963, **2**, 263–271.

Floyd, R. L. Semantic satiation: Replication and test of further implications. *Psychol. Rep.*, 1962, **11**, 274.

Helson, H. Adaptation-level as a frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

Helson, H. Current trends and issues in adaptation-level theory. *Amer. Psychologist*, 1964, **19**, 26–37.

Hull, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.

Jakobovits, L. A. Effects of repeated stimulation on cognitive aspects of behavior: Some experiments on the phenomenon of semantic satiation. Unpublished doctoral dissertation, McGill University, 1962.

Jakobovits, L. A., & Lambert, W. E. Semantic satiation among bilinguals. *J. exp. Psychol.*, 1961, **62**, 576–582.

Jakobovits, L. A., & Lambert, W. E. Mediated satiation in verbal transfer. *J. exp. Psychol.*, 1962, **64**, 346–351. (a)

Jakobovits, L. A., & Lambert, W. E. Semantic satiation in an addition task. *Canad. J. Psychol.*, 1962, **16**, 112–119. (b)

Jakobovits, L. A., & Lambert, W. E. Stimulus characteristics as determinants of semantic changes with repeated presentation. *Amer. J. Psychol.*, 1964, **77**, 84–92.

Johnson, R. C., Thomson, C. W., & Frincke, G. Word values, word frequency, and visual duration thresholds. *Psychol. Rev.*, 1960, **67**, 332–342.

Kanungo, R. N., & Lambert, W. E. Paired-associate learning as a function of stimulus and response satiation. *Brit. J. Psychol.*, 1963, **54**, 135–144. (a)

Kanungo, R. N., & Lambert, W. E. Semantic satiation and meaningfulness. *Amer. J. Psychol.*, 1963, **76**, 421–428. (b)

Kanungo, R. N., Lambert, W. E., & Mauer, S. M. Semantic satiation and paired-associate learning. *J. exp. Psychol.*, 1962, **64**, 600–607.

Lambert, W. E., & Jakobovits, L. A. Verbal satiation and changes in the intensity of meaning. *J. exp. Psychol.*, 1960, **60**, 376–383.

Mason, M. Changes in the galvanic skin response accompanying reports of changes in meaning during oral repetition. *J. gen. Psychol.*, 1941, **25**, 353–401.

Messer, S., Jakobovits, L. A., Kanungo, R. N., & Lambert, W. E. Semantic satiation of words and numbers. *Brit. J. Psychol.*, 1964, **55**, 155–163.

Miller, A. Verbal satiation and the role of concurrent activity. *J. abnorm. soc. Psychol.*, 1963, **66**, 206–212.

Osgood, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.

Osgood, C. E., Suci, G., & Tannenbaum, P. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

Podell, J. E. A comparison of generalization and adaptation level as theories of connotation. *J. abnorm. soc. Psychol.*, 1961, **62**, 593–597.

Postman, L. The experimental analysis of motivational factors in perception. In J. S. Brown (Ed.), *Current theory and research in motivation*. Lincoln: Univer. Nebraska Press, 1953. Pp. 59–108.

Severance, E., & Washburn, M. F. The loss of associative power in words after long fixation. *Amer. J. Psychol.*, 1907, **18**, 182–186.

Smith, D. E. P., & Raygor, A. L. Verbal satiation and personality. *J. abnorm. soc. Psychol.*, 1956, **52**, 323–326.

Staats, A. W., & Staats, C. K. Effect of number of trials on the language conditioning of meaning. *J. gen. Psychol.*, 1959, **61**, 211–223. (a)

Staats, A. W., & Staats, C. K. Meaning and *m*: Correlated but separate. *Psychol. Rev.*, 1959, **66**, 136–144. (b)

Warren, R. M. Illusory changes in repeated words: Differences between young adults and the aged. *Amer. J. Psychol.*, 1961, **74**, 506–516.

Wertheimer, M. The relation between the sound of a word and its meaning. *Amer. J. Psychol.*, 1958, **71**, 412–415.

Wertheimer, M., & Gillis, W. M. Satiation and the rate of lapse of verbal meaning. *J. gen. Psychol.*, 1958, **59**, 79–85.

Yelen, D. R., & Schulz, R. W. Verbal satiation? *J. verbal Learn. verbal Behav.*, 1963, **1**, 372–377.

(Received December 16, 1963)

Psychological Bulletin
1964, Vol. 62, No. 4, 287–288

SOME COMMENTS ON "CONGENITAL INSENSITIVITY TO PAIN: A CRITIQUE"

STANLEY F. SCHNEIDER
National Institute of Mental Health, Bethesda, Maryland

A note on the paper by Sternbach to correct the statement attributed to Schneider that the latter considers the syndrome of indifference to pain a defense mechanism, which, in fact, was not the case. A distinction is made between syndrome and defense. A summary of the neuropathologic findings on one of Schneider's cases is presented. Since these findings were equivocal in pathologic significance and thus inconclusive, an explanation of the anomaly on the basis of neural deficit remains as hypothetical as those based upon psychological factors. Sternbach's concluding generalization that pain is not a necessary component in normal personality development is felt to be premature, since the evidence is hardly adequate at this time to describe how the absence of pain may effect personality.

A footnote to Sternbach's (1963) recent paper on insensitivity to pain reads: "Schneider (1962) considers the syndrome a defense mechanism [p. 246]." Since I have no wish to be misquoted, and even less to be misunderstood, a correction to this footnote is in order. The reader is referred to the original article in question (Schneider, 1962) and to the subsequent exchange of letters between Sternbach and Quarton (1963) and Schneider (1963).

It is necessary to make a distinction between a syndrome or group of symptoms (in this instance, indifference to pain) and the defense mechanism or mechanisms associated with the syndrome. It was proposed that denial and repression were the defense mechanisms involved in the genesis of indifference to pain for the three cases presented (Schneider, 1962). The syndrome itself was not considered a defense mechanism, since this would blur the difference between defense and symptom, between process and result. Sternbach and Quarton (1963) also failed to make this distinction in an earlier communication.

It is Sternbach's privilege not to subscribe to a psychological explanation of indifference to pain. He has made a valuable contribution by underlining the heterogeneity of cases in the literature, and he clearly has his own explanatory construct for the syndrome, that is, that there are several types of insensitivity to pain resulting from some sort of neural deficit. Variations in the nature and/or locus of this deficit would presumably account for the heterogeneity which he noted.

Of further interest in this regard is Magee's recent report of neuropathologic findings on one of the cases discussed by Magee, Schneider, and Rosenzweig (1961) and by Schneider (1962). Magee (1963) states,

> The only possible abnormality detected in this entire series [of pathological sections] throughout the basal ganglia, internal capsule, and thalamus was slight pallor suggesting diminished myelination in the rostral portion of the subcallosal fasciculus [p. 637].

He cautiously labels this finding equivocal in its pathological significance and concludes that the neuropathologic studies

> failed to demonstrate any obvious anatomic defect to account for the clinical features of the disorder [p. 639].

He suggests that there may be undeveloped connecting fibres from the frontal cortex to the subcallosal fasciculus, but that these cannot be detected since they would form only a small group traversing the main mass of white matter in the cortex, which would obscure them. Magee correctly observes that neither of the inconclusive necropsies reported (Baxter & Olszewski, 1960; Magee, 1963) rules out the presence of some anatomic substrate which might account for the syndrome, but at this time the neural defect idea also remains a hypothesis. We have still not determined what, or how many things, may be responsible for what is observed clinically, and there is reason to encourage several approaches to the problem.

Sternbach (1963) minimizes the possibility of any psychogenetic or even psychological factors in pain indifference and he concludes his paper with a statement that seems dubious to this writer:

> As is to be expected, the ability to survive is seriously impaired [in persons with pain insensitivity] and depends in large measure on their ability to utilize other sensory cues of actual or potential tissue damage. Personality, however, is rarely affected by this sensory deficit, and it is judged that pain is not a necessary component in its normal development [p. 262].

It is difficult indeed to see how something so necessary to the ability to survive could be so lightly dismissed. It seems more appropriate to continue to examine *how* the sensory deficit may effect personality and how it may be related to abilities to use other cues to guarantee survival, rather than to focus upon the value judgment of whether the effects add up to something "normal" or "abnormal." The few well-documented cases in the literature hardly permit a firm answer to even the first question at this time.

REFERENCES

BAXTER, D. W., & OLSZEWSKI, J. Congenital universal insensitivity to pain. *Brain,* 1960, **83**, 381–393.

MAGEE, K. R. Congenital indifference to pain: An anatomicopathologic study. *AMA Arch. Neurol.,* 1963, **9**, 635–640.

MAGEE, K. R., SCHNEIDER, S. F., & ROSENZWEIG, N. Congenital indifference to pain. *J. nerv. ment. Dis.,* 1961, **132**, 249–259.

SCHNEIDER, S. F. A psychological basis for indifference to pain. *Psychosom. Med.,* 1962, **24**, 119–132.

SCHNEIDER, S. F. Letter to the editor. *Psychosom. Med.,* 1963, **25**, 85–88.

STERNBACH, R. A. Congenital insensitivity to pain: A critique. *Psychol. Bull.,* 1963, **60**, 252–264.

STERNBACH, R. A., & QUARTON, G. C. Letter to the editor. *Psychosom. Med.,* 1963, **25**, 83–85.

(Received April 2, 1964)

Vol. 62, No. 5 November 1964

Psychological Bulletin

THE PERSONALITY OF THE PERCEIVER AS A FACTOR IN PERSON PERCEPTION[1]

SID SHRAUGER and JOHN ALTROCCHI

Duke University

This review shows that research on accuracy and assumed similarity has primarily led to blind alleys thus far, but investigations of attribution of specific traits, differentiation, free descriptions, and implicit personality theories show promise. Theories which attend to personality and social factors together and research dealing with personality, situational, and stimulus variables together are needed. The authors offer a framework for future research which delineates phases of the judging process and which suggests that cognitive control variables may be especially prominent in selection of cues, motivational-interpersonal variables may be important in drawing inferences from cues, and situational variables may be salient in making a verbal report.

The factors influencing the perception of other people can be organized into three sets of variables and the interactions among them: the attributes of the stimulus person, the nature of the interaction situation, and the characteristics of the perceiver (Tagiuri, 1958a). This paper will focus on the third set of variables, conceptualized as personality variables in the perceiver which influence how he perceives others. Despite the generally held assumption that one's perceptions of others are partially shaped by his own personal characteristics, it will be seen that there has been little research devoted directly to this topic and that contributions to the area have been scattered and unsystematized. This paper will review the few directly relevant studies, abstract pertinent findings and ideas from related areas, and present a framework for future research involving personality and person perception.

We will use the term personality to refer to the relatively enduring patterning of a person's dispositions to think, feel, and behave in certain ways. Such a broad definition almost covers the area of individual differences. Following Secord (1958), however, the term will not refer to cultural determinants of consensual perception of others but will

[1] Some of the ideas expressed in this paper were briefly presented in a symposium paper by the junior author at the Southeastern Psychological Association in Louisville, March 1962. An early development of the present paper was submitted in 1963 by the senior author as a major area paper in partial fulfillment of the requirements for the PhD at Duke University.

The authors wish to express their gratitude to David A. Rodgers, Edward E. Jones, Douglas N. Jackson, Michael A. Wallach, Robert C. Carson, and Kurt W. Back for their helpful advice and criticism, and to the Scripps Clinic and Research Foundation, La Jolla, California, where parts of this paper were written while the junior author was on sabbatical leave. Both authors are associated with the Departments of Psychology and Psychiatry, Duke University.

refer to differences among perceivers within a given culture. In line with current definitions of person perception (Tagiuri, 1958a), we will use this term to refer to attribution of psychological characteristics (e.g., traits, intentions, emotions) to other people—either by describing them or by making predictions of their subsequent behavior. This broad definition includes processes of inference which have often been classified as cognition and apperception rather than as perception. Research on attribution of traits to people on projective tests such as the TAT will not be included. Although such research has some relevance for this area, the task or set in projective tests emphasizes imagination and fantasy, while the studies reviewed here involve a set to be accurate or realistic, which is presumably more relevant for understanding the ways in which people characteristically perceive others in real life situations.

Accuracy

The earliest experimental investigations of how people differ in their attribution of personal characteristics to others were concerned with accuracy. Two comprehensive reviews of research done before 1954 on accuracy of perception (Bruner & Tagiuri, 1954; Taft, 1955) make a comprehensive rereview here unnecessary. These reviews concluded that accuracy is positively correlated with intellectual and social skills and adjustment. Cronbach's (1955) incisive critique of accuracy scores, published at almost the same time, however, rendered these conclusions equivocal. Accuracy was typically measured by one of two methods, both of which have serious shortcomings. One method employed a comparison between subjects' predictions of how a stimulus person would fill out an inventory and how the person actually filled it out. This is a task which involves the contamination of response sets and the influence of the unknown factors that mediate responses to an inventory, and is also a task which is of questionable relevance unless one is specifically interested in the ability to predict others' modes of self-presentation. The second method involved a comparison between a subject's description of a stimulus person and the mean of all subjects' descriptions of that person. Without even considering the effect on the latter measure of biases held by other judges, the effect of regression to the mean is that the person with the highest accuracy score is the one who most accurately judges the average response of the group. Using either of these methods it is impossible to determine the degree to which high accuracy reflects an accurate stereotype of the group judged or an ability to sense differences between others or both.

These difficulties with criteria of accuracy have brought methodological advances. Techniques have been developed which eliminate meaningless artifacts and response sets by using comparisons between deviation scores instead of absolute scale values. Bronfenbrenner, Harding, and Gallwey (1958) have used correlations between deviation scores to compare one subject's ratings of a set of stimulus persons with the mean ratings of them by all subjects, while Carson has used a correlation of deviation scores across items to assess the accuracy of a subject's prediction of a stimulus person's self-description.[2] Measures of differential and stereotype accuracy have been conceptually and methodologically separated (Bronfenbrenner et al., 1958; Sechrest & Jackson, 1961).[3]

[2] R. C. Carson, personal communication, 1963.

[3] Satisfactory measures of other aspects of accuracy less germane to the focus of this review have also been developed—for example,

Results of these later studies suggest that the personality traits associated with stereotype and differential accuracy are not the same. A relatively consistent finding has been that women high in stereotype accuracy are well liked, well socialized, and predictable (Bronfenbrenner et al., 1958; Sechrest & Jackson, 1961). Consistent personality differences related to differential accuracy have not yet been noted, possibly because of lack of consistency in measures of differential accuracy.

The often implicit assumption that accuracy is a trait which is relatively stable over a variety of persons, characteristics, and situations has recently been put to test. The results are mixed (e.g., Cline & Richards, 1960; Crow & Hammond, 1957). What little individual consistency has been shown is attributable to stereotype accuracy. Chance and Meaders (1960) used a technique for studying accuracy which assures some generality over persons judged. Two stimulus persons were selected who were at opposite extremes on the dimensions judged; and the personality characteristics of judges, who were high in accuracy for both stimulus persons, were compared with those who were low for both stimulus persons. The results suggest that the more accurate person tends to be outgoing and sociable but little given to reflection about his interpersonal behavior. Regarding differential accuracy, however, the work of Bronfenbrenner et al. (1958) has demonstrated behavioral differences between subjects accurate in judging opposite-sexed persons and subjects accurate in judging same-sexed persons, thus clearly demonstrating one fallacy in assuming generality over stimulus persons.

Despite attempts to refine measures of accuracy, one cannot, on the basis of current research evidence, be assured that people at one extreme on any personality dimension are consistently more prone to perceive specific kinds of other people more accurately than are people at the other extreme. If one's goal is to determine how personality relates to perception of others, investigations of accuracy will have to be supplemented by, or perhaps replaced by, studies of the components of accuracy and studies of attribution of specific traits to others. Such specific investigations have emerged only recently, however, and preceding this development there was a search for another kind of generality in perception of others.

Assumed Similarity

Growing out of Rogers' (1951) self-concept theory and Fiedler's early research (reviewed in Fiedler, 1958), a number of studies of assumed similarity appeared. It was proposed that some people (e.g., successful psychotherapists, effective team leaders) assume that other people are similar to themselves and thus attribute approximately the same traits to others as to themselves. In another penetrating paper, however, Cronbach (1958) rendered many earlier conclusions equivocal by pointing out that assumed similarity scores confound many different components. In the first place, an overall assumed similarity score may combine judgments of several different trait dimensions into a single score. Furthermore, even with a single trait, similarity scores can be analyzed into five simpler mathematical components—the means and standard deviations of the two descriptions (self and other) and the re-

Tagiuri's (1958b) measures of accuracy of group members' sociometric preferences for one another and Ekman's measures of accuracy of matching verbal statements and photographs taken when the statements were being made (P. Ekman, personal communication, 1963).

gression of one on the other. Cronbach pointed out that only after the monadic effects have been given separate consideration can one speak with any certainty about the dyadic component—the regression or correlation component. A study by Altrocchi (1961) decisively supported Cronbach's conclusions by presenting differences in assumed similarity scores which were due to intergroup differences in self-description, with no evidence of differences in person perception. There is a consistent tendency for subjects to describe strangers favorably in psychological experiments and therefore anyone with a relatively favorable self-concept will seem to be assuming similarity.

The statement of research results in terms of global tendencies such as accuracy and assumed similarity may lead to confusion rather than to clarification. A case in point is a series of studies on authoritarianism. The principal conclusion of the first study was that nonauthoritarians were more accurate than authoritarians (Scodel & Mussen, 1953). Later investigations, however, suggested that the results might be more parsimoniously explained as differences in assumed similarity (Crockett & Meidinger, 1956; Scodel & Friedman, 1956), but a subsequent study showed that these differences in assumed similarity were not consistent from one instrument to another (Schulberg, 1961). While the authoritarian and nonauthoritarian may not differ in their tendencies to be accurate or to assume similarity, these and other studies have consistently found that people who are high in authoritarianism tend to see others as more authoritarian than do nonauthoritarians. These results point to the utility of a strategy in which individual differences in the attribution of particular traits are attended to in themselves rather than being interpreted in terms of global variables such as accuracy and assumed similarity.

Differentiation among Others

Cronbach's (1955) methodological suggestions and other recent contributors to personality theory and research have emphasized the importance of the extent to which one forms a differentiated conceptualization of his social environment (Bieri, 1961; Harvey, Hunt & Schroder, 1961; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962). Conceptually, differentiation refers to the tendency to make fine distinctions among people and thus to perceive them as different from one another. Operationally, it has been defined in several ways, and little is known yet about the interrelationships among these measures. It has often been defined as the standard deviation of one's judgment of a number of people on a specified dimension. When measured this way, differentiation has not been shown to be consistently related to personality traits, despite the fact that this measure is somewhat consistent over dimensions judged (Rabin, 1962). Fiedler (1958) reported no significant correlations between subjects' scores on a number of personality tests and their tendencies to differentiate in their descriptions of people best and least liked, while Koltuv (1960) found no relationship between subjects' tendencies to differentiate and ratings of the social favorability of their behavior made by peers. Ormont (1960) found no relationships between group therapy patients' scores on Welsh's Anxiety scale and measures of their tendencies to differentiate in their *Q* sort descriptions of other group therapy patients.

Measures of adjustment and the tendency to differentiate have, however, been shown to be related in two studies. Ormont (1960) reported that group ther-

apy patients who were classed as better adjusted tended to differentiate more in their descriptions of the other members of their therapy group than did patients who were more poorly adjusted. Rabin (1962) analyzed descriptions of diverse stimulus persons by adjusted and maladjusted subjects, using criteria for maladjustment which included whether or not subjects had ever sought help for emotional problems. Maladjusted subjects had larger standard deviations in their judgments on each of four factorially determined dimensions of judgment. Such a result could be a function of increased random responding of maladjusted subjects.[4] Numerous procedural differences between the two studies might account for the seemingly discrepant results. These results, however, can be seen as supporting the plausible proposition that differentiation increases and then decreases as one moves from extreme repressiveness to awareness of personal shortcomings in self and others to severe disruption of personality functioning. In support of such a proposition, Altrocchi (1961) demonstrated a slight tendency among normals for sensitizers to differentiate among others more than repressors and he suggested pursuit of this problem because of its relationship to research demonstrating more general perceptual differences (e.g., Holzman & Gardner's, 1959, demonstration that extreme repressors tend to be levelers on a neutral psychophysical task).

One problem in using a variance measure as an index of differentiation and relating it to personality characteristics of judges is that high scores might be achieved by two psychologically different processes. High scores would be obtained just as readily by using only the extremes of rating scales and classifying people into two distinct types as by making fine distinctions and distributing people along the entire continuum. Making extreme ratings and forming dichotomized judgments of others may be related to traits such as impulsivity and intolerance for ambiguity, while making fine distinctions among others and using a large number of categories for evaluation may be more reflective of social sensitivity. Thus, use of a variance measure of differentiation requires attention to the exact contributions to high variance scores. New and refined dimensional procedures (Jackson, 1962; Jackson & Messick, 1963) will probably, in fact, prove much more adequate than the variance model for investigating differentiation.

Other operational measures of differentiation seem to define it more specifically as the tendency to emphasize differences among others and to see people as distinctly different from each other. There have, for instance, been a number of investigations of cognitive complexity, which is defined as the number of independent dimensions which people characteristically utilize in describing others on a modified Kelly Role Construct Repertory Test (Kelly, 1955). A more differentiated conceptual system with a greater number of descriptive dimensions available would presumably allow for a more precise, unique description of other people. People having a more differentiated or cognitively complex conceptual system have been shown to give more complex Rorschach percepts, to be better able to predict how others will respond in a series of social situations, and to be less likely to change their initial impressions of others after receiving contradictory information about them (cf. review by Bieri, 1961). Gardner and Schoen (1962) have criticized measures of cog-

[4] D. N. Jackson, personal communication, 1964.

nitive complexity as too indirect and too much influenced by verbal fluency and have shown that cognitive complexity is unrelated to their own measure of differentiation. Their measure, called conceptual differentiation, refers to the number of categories into which heterogeneous stimulus objects are grouped. They have proposed that conceptual differentiation may optimally be viewed as a basic characteristic of cognitive organization. Research is now in progress at the Menninger Foundation to determine the relationship between conceptual differentiation and measures of differentiation in social perception.[5] The relationships between the different measures of differentiation and cognitive complexity have yet to be clarified, but the present authors think that research in this area offers considerable promise for delineating individual differences in the perception of people.

Judging on Specified Descriptive Dimensions

Some studies scattered throughout the psychological literature of the last 15 years have focused on differences in how others are characteristically described on dimensions specified by the experimenter. They can be grouped under four headings: favorability, hostility, dominance, and interactions between the traits of the judge and the traits of the stimulus person.

Favorability

The most prominent feature of any description of a person is the degree of favorability expressed. Factor analyses have consistently indicated that an evaluative factor accounts for the largest portion of variance in descriptions (Levy & Dugan, 1960; Osgood, Suci, & Tannenbaum, 1957). Regarding individual differences in general favorability toward others, maladjusted subjects (using a loose Rorschach criterion) have been shown to rate the "real" but not the "apparent" personality of the stimulus person less favorably than well-adjusted subjects (Matkom, 1963). Edwards (1959) showed that people who are more positive in their evaluation of others are also more positive in their self-descriptions. Studies testing Rogers' (1951) notion, that the person who is accepting of himself is also likely to be more accepting of other people, have shown generally positive results as indicated by a variety of questionnaires exploring attitudes toward the self and others (e.g., Berger, 1952; Omwake, 1954; Phillips, 1951; Suinn, 1961). Some of these findings, however, may readily be interpreted in terms of a response set specific to a certain measuring instrument. Furthermore, there has not been satisfactory support for Rogers' hypothesis that an increase in one's self-acceptance is a causal determinant of increased acceptance of others. It should also be pointed out that acceptance of others is not the same as seeing them favorably. On the contrary, Rogers (1951) noted that the self-accepting person will more readily recognize negative aspects of others when this is justified, since he will not be threatened and will not distort his perceptions in order to defend himself. Thus it is important to separate acceptance of others from esteem of others or favorability of one's description of others, just as it is important to separate self-acceptance from self-esteem (Hatfield, 1958). Finally, investigation of favorability needs to be focused on differences in favorability of perceptions of particular types of stimulus persons as well as on the general tendency to describe others favorably.

Favorability of description of others

[5] R. W. Gardner, personal communication, 1963.

should probably be high on a priority list of specific descriptive dimensions worth investigating since it is a prominent variable in descriptions of people and has strong implications for the type of interaction which one might be expected to have with the persons described (Tagiuri, 1958b), especially if subjects are describing the real instead of the apparent personality of the stimulus person. Matkom's (1963) subjects were evidently able to make this distinction easily, and it might lessen subjects' tendency to say only blandly favorable things about others. While there have been some careful investigations of situational factors and attributes of the stimulus person which lead to favorable perceptions (e.g., Jones, Gergen, & Jones, 1963), there has been little work on the effects of perceiver variables and their interactions with these other factors (cf. Jones & Daugherty, 1959).

Hostility

Leary (1957) reported that people who described themselves and were described by others as hostile tended to attribute considerable hostility to others, while people described as exceptionally friendly tended not to attribute hostility to others. Berkowitz (1959) showed that anti-Semitic students, after being frustrated by a hostile experimenter, tended to generalize their feelings and rate others as more hostile than did students who scored low on anti-Semitism. Cohen (1956) found that college males utilizing projection as their principal defense, after discussing with partners a story related to some area of psychosexual conflict, described their partners as more hostile when the partners were also projectors than did subjects using other defensive processes. Because subjects and stimulus persons interacted in all these studies, however, one cannot be sure whether differences in the attribution of hostility were due to individual differences in the interpretation of essentially similar behavior or to differential influence of the subjects on the persons described. Cohen's experiment, nevertheless, stands out in this research area thus far because of his theoretically derived attention to individual differences in subjects and stimulus persons and specification of the situation—in this case the presumed arousal of threat in the subjects.

Other studies involving hostility neglected threat arousal but were based on clinical theories and research in perception and personality (Eriksen, 1963) which suggest that the repressive person tends to fail to perceive negative attributes and hostile actions of others while certain other people are especially sensitive to such attributes. Altrocchi and Perlitsh (1963) compared the descriptions of classmates by student nurses who were classified on the basis of their use of characteristic defensive patterns. They found that people classed as expressors, individuals characterized by their expression of impulses and lack of anxiety over this expression, were more likely to attribute hostility to others than were either sensitizers, people with high anxiety who rely heavily on obsessive and intellectualizing defenses, or repressors, who primarily utilize defenses of denial and repression. A second study (Altrocchi, Shrauger, & McLeod, 1964) provided more experimental control but failed to confirm the earlier findings and found different results for male and female subjects. On the other hand, McDonald (1965), using the same instruments as Altrocchi and Perlitsh (1963) and studying 177 single pregnant females' descriptions of self and parents, supported the Altrocchi and Perlitsh and the Altrocchi, Shrauger, and McLeod predictions in a convincing fashion. Again, a primary

task in clarifying such results is separating differences in perception from differential influence (provoked hostility). Secondly, there are many reasons to believe that there may be major sex differences in the ramifications of various patterns of ego control in personality functioning and in the meaning of attribution of hostility, dominance, and other traits. Furthermore, experimental manipulation of threat may be particularly important when classifying subjects by patterns of ego defense, since some of the differences in perception of others predicted by clinical theories may apply only in threatening situations which arouse ego defenses. Feshbach and his collaborators have been quite successful in arousing hostility (which would theoretically be psychodynamically threatening to some subjects) by disparaging subjects and have demonstrated that under these conditions hostility is more likely to be attributed to persons similar than to persons dissimilar to the subjects (Feshbach, Singer, & Feshbach, 1963). Inclusion of perceiver personality variables would seem to be a useful addition to such research and Feshbach's techniques of arousal of hostility would seem to be a useful addition to studies of personality factors in perception of hostility in others.

Dominance

Early studies on attribution of dominance confront one again with the alternative explanations of differences in perception versus differential influence. Naboisek (1953) and Leary (1957) reported that autocratic, exploitive, dominant people described others as weaker than they were described to be by most people. Altrocchi (1959) and Smelser (1961), however, reported controlled studies in which there were no such differences in attribution of dominance. In Altrocchi's study the subjects watched a movie of the stimulus persons in action, but in Smelser's study the subjects and stimulus persons interacted with each other. While Smelser's (1961) report concludes that "The degree of dominance ascribed to the partner, relative to the subject's own dominance, is a function of the subject's personal dominance [p. 540]," it is apparent that his was a dyadic frame of reference.[6] Examination of his results shows no complementary attribution of dominance when attribution of dominance to others is kept separate from attribution of dominance to self. These results at least limit the generality of Naboisek's and Leary's observations and suggest that their results may have been due to stronger affect arousal in the clinical settings they were using, or to differential influence, or both.

A recent paper by Bieri (1962) has increased the sophistication of research on attribution of dominance and hostility. He found that female subjects' ratings of a male stimulus person on dominance were a linear function of the relative frequency of dominant and submissive behaviors of the stimulus person in the information provided by the experimenter, but male subjects' ratings followed what appeared to be a stepwise function. The results for the love-hostility dimension were somewhat different. These results suggest that differences in attribution of traits are a function of sex-related personality variables in interaction with the type of behavior judged and varying degrees of inconsistency in the stimulus behavior.

Interactions between Traits of the Judge and Traits of the Stimulus Person

While we have been considering differences in the descriptions of people in

[6] W. T. Smelser, personal communication, 1963.

general, some differences in the attribution of traits to others are evident only when particular kinds of other people are described. For example, men's perceptions of their mothers have been shown to influence the sorts of characteristics which they ascribed to other women, but only for women described as motherly (Secord & Jourard, 1956). It is implicit in the definition of certain personality traits such as racial prejudice or authoritarianism that the trait will only affect one's perception of certain types of people. The effect of one's authoritarianism upon his judgment of others is dependent on the power or status of the person judged. If the person judged is presented as a leader with high status, the authoritarian describes him in more generally favorable terms than the nonauthoritarian (Jones, 1954) and is less likely to lower his opinion of him even when he is later shown to be an unpleasant person and a poor leader (Thibaut & Reicken, 1955). Also, authoritarian college students tend to describe others presented as college students or peers more favorably than do nonauthoritarians (Kates, 1959), but when those judged are presented as strangers not particularly similar to the judges, it is the nonauthoritarians whose descriptions are more favorable (DeSoto, Kuethe, & Wunderlich, 1960). With such evidence available, it would seem that investigators should use stimulus persons specifically chosen to represent groups the subjects will perceive differently or should use a broad and representative sample of stimulus persons so that specific interactions will cancel each other out.

Judging without Specification of Descriptive Dimensions

Studies in the preceding section have been concerned with differences in how people judge others along dimensions or on traits which the experimenter has specified. There is no guarantee, however, that the dimensions in which the experimenter is interested are also those which people commonly utilize or can utilize in judging others. Much of the individual variability in descriptions may well be due to differences in the types of dimensions or traits which people spontaneously use in judging others. There have been a number of pleas for research to determine the salient dimensions typically used in evaluating others and to delineate how people differ in the dimensions which they employ (Beach & Wertheimer, 1961; Bruner & Tagiuri, 1954; Cronbach, 1958; Hastorf, Richardson, & Dornbusch, 1958; Tagiuri, 1958a). While it has been shown that own and experimenter-specified traits provide reasonably comparable measures of cognitive complexity (Tripodi & Bieri, 1963), studies investigating subjects' own spontaneous descriptions have also been shown to offer unique information about individual differences in person perception.

Early and primary empirical attention to analysis of subjects' descriptions focused on the use of psychological as opposed to physical characteristics (Sarbin, 1954). The tendency to describe others in external-physical in contrast to internal-psychological terms has been interpreted as one aspect of the characteristic of field dependency (cf. review by Bieri, 1961) and has also been positively related to authoritarianism (Wilkins & deCharms, 1962). There is some evidence that people show consistent differences in whether they select psychological or physical characteristics (Wolin, 1956), and there are sex differences in this respect. When classifying individuals into groups and when listing traits describing them, men have been found to focus more on descriptive

characteristics such as the color or types of clothing worn, height, and direction faced; while women have been found to be more prone to focus on inferential, psychological characteristics (Sarbin, 1954). Sex differences occur when extensive free descriptions are analyzed, but the descriptive versus inferential dichotomy seems to be an oversimplification (Beach & Wertheimer, 1961). The differences seem partially dependent on the sorts of people described and on the differential importance of the specific behaviors in the male and female role. Studies with children indicate that boys tend to emphasize behavior associated with aggression, nonconformity, and physical recreation; while girls focus more on nurturant behavior, physical appearance, and social skills (Campbell & Radke-Yarrow, 1956; Hastorf, 1962).

A recent investigation by Hastorf (1962), however, suggested that the matter is even more complicated. He demonstrated that if behavior relevant to a particular personality dimension is mentioned frequently in descriptions, the describer may be having trouble in his own adjustment with respect to that dimension. For example, children who are more dependent tend to emphasize information on generosity, giving aid, and needing aid in their descriptions. Boys who are not aggressive tend to mention aggressive behavior more often in their descriptions of others than do more aggressive boys. With girls, the differences are opposite; those who are rated as more aggressive tend to mention aggression more in describing others than do those who are not so aggressive. It would be hard to obtain such information if subjects were not allowed some freedom in describing others. Thus the suggestions that subjects be allowed to describe others in their own terms clearly have some merit.

Implicit Personality Theories

There is one further historical trend, and it comes last in this review because of the sophistication of some of the studies, even though it includes some of the earliest forays into personality factors in person perception (Secord & Muthard, 1955). This is the notion of implicit personality theory, first introduced by Bruner and Tagiuri (1954, p. 649), elaborated and named by Cronbach (1955), and centrally important in Kelly's (1955) theory of personality. Cronbach suggested that perceivers differ in: response biases toward rating consistently higher (or lower) on particular traits, tendencies to make more extreme (or more central) ratings on certain traits, and tendencies to associate particular traits with each other.

Regarding the first tendency, previous sections of this paper have shown that under certain conditions different personality groups tend to rate others higher or lower on particular traits (e.g., hostility), although this may well be a function of the particular types of people described or the differential nature of the interaction with those described. Gross (1961) showed that consistent individual differences in the tendency to rate others as high or low on particular traits are small when a heterogeneous set of stimulus persons are judged. She concluded that the concept of a unitary attitude toward the generalized other is too nonspecific to be useful, so that such terms as implicit personality theory would be more precise if used in the plural. Thus she suggested that the contribution of personality variables can be seen most clearly in conjunction with specific situations and specified stimulus persons—an interactional recommendation which is a cornerstone of this review.

The tendency to make more extreme

ratings has been mentioned in the section on differentiation. There have, however, been no investigations of differences in the extent to which stimulus persons are distributed on relevant as opposed to nonrelevant dimensions or of personality-related differences in the extremeness with which others are perceived on particular dimensions.

The third aspect of implicit personality theory, patterns of perceived consistency in trait intercorrelations, has begun to receive exceptionally competent research attention. Koltuv (1962) contributed a scholarly historical review of trait centrality and trait intercorrelations, which renders a review unnecessary here. She suggested that lack of attention to the relevance of the traits for each individual may account for Gross's (1961) finding that perceiver predispositions account for little of the variance in person perception. Koltuv obtained the names of personally relevant and nonrelevant traits and the names of familiar and unfamiliar people for each of her subjects and then had them rate these people on both kinds of traits. Her results strongly indicated the importance of the correlational aspect of implicit personality theories by demonstrating that: each perceiver assumed a matrix of correlations to exist among traits, these correlations were greater for unfamiliar than familiar persons, correlations were greater among personally relevant than among nonrelevant traits, and these intercorrelations were not wholly a function of the overall evaluative attitude of the subject (halo effect) or the logical connections between certain traits. She concluded that idiographic analysis is necessary to demonstrate the magnitude of individual intercorrelations which otherwise might be obscured by nomothetic analysis.

A study by Secord and Berscheid (1963) also demonstrated the generality of trait intercorrelations, showing that certain types of traits are perceived as highly intercorrelated not only in judgments of people in whom these traits are commonly thought to appear together, but in other people as well. Both Koltuv (1962) and Secord (1958) pointed out that it is unrealistic to assume that people apply their implicit personality theories indiscriminately in the judgment of all people. Having established the existence of consistent patterns of trait intercorrelations, it will be useful to investigate the determinants of differences in these patterns. Theoretical models and statistical methods which can encompass the complexity of such multidimensional problems as analyzing different patterns of trait relationships for different kinds of subjects, situations, and stimulus persons have been explicated and are available (Jackson, 1962; Jackson & Messick, 1963; Tucker, 1963). It is likely that use of such methods will increase our understanding of person perception considerably.

A Framework for Future Research

The diversity of contributions, lack of organization, and variable usefulness of research dealing with personality variables in person perception is apparent. In an attempt to provide some organization for the area and in order to suggest some potentially productive approaches for future research, an organized framework will be presented. It includes suggested personality variables, some possible interactions between personality, situational, and perceptual variables, a delineation of phases of the judging process, and some general methodological recommendations.

Personality Variables

Neither personality theories nor empirical research thus far have clearly

spotlighted the independent personality variables that might be most significantly related to person perception, but modern developments of interpersonal theory and perceptual-cognitive research based on modified psychoanalytic theory suggest the usefulness of two kinds of personality variables: multilevel interpersonal motives or traits and cognitive controls.

Freud's (1924) conceptualization of transference as a source of perceptual distortion in the psychoanalytic situation, the extension of the concept by Fenichel (1945) and Sullivan (1948) to all types of interpersonal situations, and Sullivan's (1948) conceptualization of parataxic distortion of and selective inattention to the anxiety-producing attributes of others led to Leary's (1957) notions concerning the effect of personality on person perception. His and subsequent similar circumplex models (Foa, 1961; Lorr & McNair, 1963) suggest the usefulness in interpersonal research of such personality variables as hostility, dominance, sociability, and combinations among them. Furthermore, Leary's organization of these variables in a conception of levels of personality provides an organized way of studying the dynamic role of possession of a trait by the perceiver and thus provides a rationale and methodology for research of the kind reported by Hastorf (1962)—although Leary's measures of the various levels leave much to be desired. Thus, for example, a person with considerable fantasy hostility who is seen by others as frequently hostile, but who does not conceive of himself as hostile, might be expected to attribute considerable hostility to others, especially under conditions involving strong affect such as anger or anxiety. Many similar hypotheses can be deduced.

At the same time, we strongly concur with Gardner and his associates (Gardner, Holzman, Klein, Linton, & Spence, 1959, and many other publications) and Jackson (1962) that consideration of motivational variables is not sufficient and should be supplemented by attention to organizational dispositions of the individual, such as cognitive controls. For instance, as mentioned in a preceding section, by Jackson (1962), and Gardner and Schoen (1962), conceptual differentiation may relate to differentiation among others in person perception. It would be eminently sensible to coordinate research in individual differences in person perception with research concerning individual consistencies in cognitive or perceptual styles by use of such cognitive control principles as conceptual differentiation (formerly equivalence range), leveling-sharpening, constricted-flexible control, scanning, and field articulation (Gardner et al., 1959; Messick & Fritzky, 1963; Witkin et al., 1962). Separate assessment of cognitive controls and multilevel interpersonal traits would probably clarify some of the results obtained using complex combinations of cognitive and motivational variables such as authoritarianism and repressiveness.

Personality-Situation Interactions

Here is where a theoretical vacuum is most apparent. Our current theories deal with personality variables or social (situational) variables but almost never both. Nevertheless, it is likely that some situational variables and some personality variables interact with each other in their effects on person perception. For instance, it is likely that subjects will differ considerably in their projection of experimentally aroused anxiety (Singer & Feshbach, 1962) and, as implied in earlier sections of this review, a subject who does not admit hostility which is perceived by others in his behavior and which is revealed in his fan-

tasy productions should be more likely to project experimentally aroused hostile affect (Feshbach et al., 1963). Furthermore, this suggests an important point, usually implicit but rarely explicit in personality theories (e.g., Leary, 1957): The presence of strong emotion may well be a necessary condition for the appearance of the effects of motivational (not cognitive) variables on person perception. As Secord, Backman, and Meredith (1962) have suggested, the static notion that perceiver traits will have a simple relation to attribution of traits to others is weak without attention to dynamic affective and role relations between the people involved. In our necessary attempts to provide rigor in the laboratory, it is easy to forget that real interpersonal encounters often involve impressive feelings of affection, hostility, and fear. We must utilize ways of arousing such feelings in our experiments or we must pay more research attention to nonlaboratory situations where emotions are naturally strong.

Person Perception Variables and Their Interactions

Concerning dependent or person perception variables, we have argued for the use of free descriptions, attention to the real as opposed to the apparent personality of the stimulus person, attention to specific traits rather than general tendencies such as assumed similarity, have stressed the importance of subjects' conceptions of intercorrelations among traits, and have pointed out that certain differences between subjects will occur only in the perception of specified stimulus persons. A new contribution to the delineation of useful dependent variables is Hastorf's current research on the perception and evaluation of behavior change in a stimulus person.[7] More generally, developments in interpersonal theory (Foa, 1961; Leary, 1957; Lorr & McNair, 1963) suggest the usefulness of such variables as hostility, dominance, and sociability as dependent as well as independent variables (cf. Bieri, 1962). The more such traits can be used in conjunction with specific interpersonal situations, the better. Finally, triple interaction studies are quite feasible—what kinds of people describing what kinds of other people in what situations—and will certainly be a necessary future step, although studies even approaching assessment of triple interaction are rare (cf. Cohen, 1956; Jones & Daugherty, 1959).

All of this leaves the perceptual-cognitive process rather global and unitary, however. Delineation of phases of the judging process may be useful to the general investigation of person perception and specifically useful in the present context by indicating how personality variables may contribute differently in the different phases. While the process of forming an impression may be perceived as an immediate, unitary phenomenon, this process can be usefully conceptualized as consisting of three phases: selecting cues, drawing inferences about personal characteristics from these cues, and translating one's impressions into an overt verbal response. The distinction between the first two phases is similar to distinctions made by Jones and Thibaut (1958), Secord (1958), and Bieri (1962).

Selection of Cues. A major factor limiting the ability to generalize to person perception from research on perception of ambiguous stimuli (Eriksen, 1963) is the complexity of a person as a stimulus object. Even a photograph presents so much sensory data that it is extremely difficult to assimilate all the

[7] A. H. Hastorf, personal communication, 1963.

cues presented. Therefore, an important source of individual variability in the judgment process may well be differences in cues chosen as bases for judgments. Selection can occur in studies in which the person judged is not actually observed by the judges, but the data about him are presented in the form of brief vignettes or descriptions of his personal attributes. One could investigate selection in this situation by analyzing what information is recalled when the stimulus information is removed (Carlson, 1961). Culture places selective emphasis on certain cues (Secord, 1958), but there may well be consistent individual differences in the degree to which particular classes of cues such as tone of voice, movements of hands, content of verbalizations, or particular physiognomic features, for example, are utilized as determinants of descriptions. If such differences were demonstrated, they might well affect the psychological attributes ascribed to others, since particular psychological characteristics tend to be inferred from particular physiognomic traits (e.g., Secord, Dukes, & Bevan, 1954). A recent exploratory foray in this area made by Exline (1963) showed that women focused visually on those with whom they interacted more than men did. This suggests a greater reliance by women on visual cues about the attributes of others.

In an ingenious series of studies, Levy (1961a, 1961b, 1963) has demonstrated the potency of perceiver attributes and the relevance of adaptation, anchoring, dissipation, conditioning, and generalization in person perception and has proposed that the learning of dispositional responses is one mechanism by means of which many of the reported effects of personality variables might be produced. He conceptualized that dispositional responses do not alter what the person sees (selection of cues), but alter the probabilities of how he interprets what he reports seeing (inference). From the arguments presented by Levy (1963) and Secord (1958) and from the evidence presented in this and the following section, one would suspect that the kinds of personality variables most likely to affect selection of cues are cognitive controls or styles, while motivational variables are more likely to affect inference. For some readers it may clarify communication to note that we are asserting that attribution of traits is probably influenced by secondary process as well as primary process variables. Furthermore, the two kinds of variables may interact. For instance, Klein (1954) suggested that needs influence cognition less in people with flexible, as opposed to constricted, cognitive control.

Inference. The most basic form of inference occurs when some cue such as an inflection of the voice, a gesture, or an eye movement is interpreted as signifying some psychological feature of the person being described. For purposes of clarification, we shall refer to this as interpretive inference. The same cue, while equally attended to, may be seen by various people as implying very different characteristics or it may be seen as reflecting nothing. One must then consider how many and what dimensions people typically have at their disposal for conceptualizing others' behavior. The number of dimensions may relate to such cognitive controls as conceptual differentiation. Descriptive dimensions must then be classified in some way. The most frequently used classification of dimensions is the dichotomous distinction between physical, external attributes as opposed to internal, psychological characteristics, although it has been shown that more discriminating categorizations can be successfully utilized (e.g., Beach & Wertheimer,

1961; Hastorf, 1962). The type of classification to be used is usually and most easily determined by the investigator's theoretical preference, but it will probably be useful to pay some attention to subjects' own classifications too.

If a particular sample of behavior is seen by a subject as relevant to some dimension of behavior, three factors may determine where the person is placed along the dimension in question. The first is the intensity of the behavior (Bieri, 1962). The second involves temporal extension (Secord, 1958)—one's willingness to generalize broadly on the basis of limited information, to assume that action in a particular situation is characteristic of the individual's typical behavior in many situations. Although we have suggested that cognitive controls play a primary role in selection of cues, they may also play a role in interpretive inference by influencing temporal extension and associated factors such as stereotyping (Secord & Berscheid, 1963), the halo effect, and overestimations of consistency (Ichheiser, 1949). Motivational variables probably play a role here too, especially when the motivational variable is the same as the dimension being judged. The third factor relevant to interpretive inference is the nature of the comparison standard against which particular cues are evaluated—people's varied norms and expectations about behavior in a particular situation. In assessing the contribution of this third factor, substantive and methodological influences may become highly entangled. It is necessary to determine whether consistent differences in ratings on a dimension are artifacts, only reflecting differential use of the measuring instrument, or are differences having substantive bases. The problem of response sets is complex, but a first and simple step would simply be to see whether the same differences occur using parallel measuring instruments.

Inference also takes place when a subject is asked to rate a person on attributes about which he has no direct information (cf. Bieri, 1962). We will refer to this as extended inference, and it includes subjects' implicit personality theories. One would expect that most personality differences in trait interrelations would be found in magnitude rather than direction of ratings because of the evaluative connotations which most traits carry and a tendency to see others as consistently favorable or unfavorable. There has, however, been a demonstration of directional differences in inference patterns, although the statistical significance of these differences was not determined. Jones (1954) found that persons presented as more democratic were judged by high authoritarians to be more undependable, rebellious, nonintrospective, harder to figure out, and more prone to act without thinking than persons presented as autocratic, while nonauthoritarians associated those characteristics more with the autocratic person. Considerable individual variability has also been shown in the types of attributes ascribed to others when seemingly inconsistent information about the person to be described has been presented. The formation of integrated impressions from discrepant information is found more frequently with females, children with higher intelligence, middle-class as opposed to lower-class children (Gollin, 1958), and children rated by their peers as being higher in social effectiveness (Campbell & Yarrow, 1961). Steiner and Johnson (1963) have shown that authoritarians tend not to alter their favorable impressions of an individual when they receive moderately derogatory information about him.

One study (Benedetti & Hill, 1960)

suggested that subjects' own possession of a trait influenced their patterns of inferences about people varying in their possession of that trait. More extensive investigation indicated that these relationships existed only for certain traits and differed for males and females (Benedetti, Morgan, & Bessemer, 1959). Secord et al. (1962) found no relationship between strength of a trait in a perceiver and the effect of that trait in the stimulus person on the pattern of inferences concerning other traits when liking and favorability were controlled. Their conclusions, however, from an otherwise well-designed study, rely entirely on the validity of the not yet validated Edwards Personal Preference Schedule (EPPS). Even if the EPPS were of demonstrated validity, though, such a one-level measure would not indicate the dynamic role of a trait in the personality (Leary, 1957). Much work remains to be done on individual differences in extended inference patterns. The recent availability of multidimensional methods (Jackson, 1962; Jackson & Messick, 1963; Tucker, 1963) makes such research much more feasible. We predict that motivational interpersonal variables will be more significant than cognitive controls in such research.

In most experimental studies of impression formation one would find it hard to decide whether differences in judgment were due primarily to individual differences in selection or inference. They can, however, be unconfounded to permit separate investigation. Selection can be estimated by determining which cues, from a large amount of information about the other person, the judge could later remember. The judge would be required to interpret the information as little as possible and only to reproduce it (cf. Carlson, 1961). In investigations of inference processes, on the other hand, one could standardize stimuli and could condition such sets as the reversal response set (Levy, 1963), or one could present only a limited amount of information pertinent to the judgment of one attribute and as neutral as possible (for people in general) with respect to those other attributes about which inferences are to be made. Thus, the least contaminated tests of the inference process would employ techniques similar to those used by Asch (1946) in which the information given consisted only of sets of traits. Studies of individual personality theories (Jackson & Messick, 1963; Koltuv, 1962), however, demonstrate the usefulness of analysis of individual rather than group results, and such analyses would clearly open the way for investigation of personality variables (Jackson, 1962).

Verbal Report. The final phase of the judgment process is making some verbal report of one's impressions. Individual variation in verbal report is shown in the candidness and completeness of one's report of his subjective impressions. The most likely lack of candor is a tendency to give a more favorable description than one actually perceived. Again, it is difficult to separate verbal report from selection and inference, but this may be done by analyzing sequences of responses (Forrest & Lee, 1962), by using projective devices and ratings of the subjects' nonverbal behavior, by investigating response variation as a function of reward and punishment (Eriksen, 1963), and by manipulating the experimental conditions. One could, for instance, compare descriptions obtained under conditions varying in the extent to which they aroused a social-desirability response set. One would expect that variance in descriptions due to differences in reporting would be associated more with situational factors operating at the time the report is made than with personality variables, but this,

like many problems suggested in this review, is an empirical problem about which we currently know almost nothing.

General Methodological Recommendations

A few general recommendations may warrant emphasis:

1. Differences in perception of others need to be separated from differences due to the effect of the perceiver as a social stimulus. This can be done by having subjects rate people they observe and also people with whom they interact under otherwise similar conditions.

2. Many of the research problems involving personality factors in the perception of people are problems in response sets, so that general response sets need to be investigated rather than eliminated. The response sets that need to be controlled or eliminated are those that are specific to a measuring instrument.

3. Evidence repeatedly shows that sex differences are crucial in this area and must be taken into account in research and in theory.

4. This review has not focused on the mode of presentation of the stimulus material, but this may be an important variable. For instance, Secord and Berscheid (1963) have suggested that making inferences about verbal traits from verbal cue traits is less susceptible to perceiver affect than is rating photographs.

5. We concur with Jackson and Messick (1963) that, for an accurate and full understanding of person perception, research methods and measurements must faithfully represent the complexity of the area—and, as this review has shown, the complexity is considerable.

Conclusions

Although focusing on personality variables, a major message of this review is that progress will most efficiently occur with attention to specific problems involving interactions among personality, situational, and perception variables. This is a formidable task, but surely not too formidable for psychologists in the age of computer technology.

Increased understanding of personality variables as they affect person perception can provide a link integrating intraindividual and interpersonal processes. Subtleties of the impression process itself may be illuminated, consistencies and seeming inconsistencies in overt behavior may be explained, and we should gain considerable understanding of how self-concepts and personality dispositions may be supported and maintained by selective attention to specific cues and idiosyncratic patterns of inference about others. To adapt a phrase still familiar from our recent history, it may be that one of the best ways in which processes of person perception are revealed is through analyzing the way different types of people perceive others and one of the best ways in which personality is revealed is through perception of others.

References

Altrocchi, J. Dominance as a factor in interpersonal choice and perception. *J. abnorm. soc. Psychol.*, 1959, **59**, 303–308.

Altrocchi, J. Interpersonal perceptions of repressors and sensitizers and component analysis of assumed similarity scores. *J. abnorm. soc. Psychol.*, 1961, **62**, 528–534.

Altrocchi, J., & Perlitsh, Hilda D. Ego control patterns and attribution of hostility. *Psychol. Rep.*, 1963, **12**, 811–818.

Altrocchi, J., Shrauger, S., & McLeod, Mary Ann. Attribution of hostility to self and others by expressors, sensitizers and repressors. *J. clin. Psychol.*, 1964, **20**, 233.

Asch, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258–290.

Beach, L., & Wertheimer, M. A free response approach to the study of person cognition. *J. abnorm. soc. Psychol.*, 1961, **62**, 367–374.

Benedetti, D. T., & Hill, J. G. A determiner of the centrality of a trait in impression formation. *J. abnorm. soc. Psychol.,* 1960, **60**, 278–280.

Benedetti, D. T., Morgan, R., & Bessemer, D. Relationships among the traits of the perceiver and those of the perceived in impression formation. Final report, September 1959, Contract AF 49(638)-33 (Phase 4), Department of the Air Force.

Berger, E. M. The relation between expressed acceptance of self and expressed acceptance of others. *J. abnorm. soc. Psychol.,* 1952, **47**, 778–782.

Berkowitz, L. Anti-Semitism and the displacement of aggression. *J. abnorm. soc. Psychol.,* 1959, **59**, 182–187.

Bieri, J. Complexity-simplicity as a personality variable in cognitive and preferential behavior. In D. W. Fiske & S. R. Maddi (Eds.), *Functions of varied experience.* Homewood, Ill.: Dorsey Press, 1961. Ch. 12.

Bieri, J. Analyzing stimulus information in social judgments. In S. Messick & J. Ross (Eds.), *Measurement in personality and cognition.* New York: Wiley, 1962.

Bronfenbrenner, U., Harding, J., & Gallwey, Mary. The measurement of skill in social perception. In D. McClelland, A. Baldwin, U. Bronfenbrenner, & F. Strodtbeck (Eds.), *Talent and society.* Princeton, N. J.: Van Nostrand, 1958.

Bruner, J. S., & Tagiuri, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. 2. Cambridge, Mass.: Addison-Wesley, 1954. Ch. 17.

Campbell, J. D., & Radke-Yarrow, Marian. Interpersonal perception and behavior in children. *Amer. Psychologist,* 1956, **11**, 416. (Abstract)

Campbell, J. D., & Yarrow, Marian R. Perceptual and behavioral correlates of social effectiveness. *Sociometry,* 1961, **24**, 1–20.

Carlson, E. R. Motivation and set in acquiring information about persons. *J. Pers.,* 1961, **29**, 285–293.

Chance, June E., & Meaders, W. Needs and interpersonal perception. *J. Pers.,* 1960, **28**, 200–209.

Cline, V. B., & Richards, J. M., Jr. Accuracy of interpersonal perception: A general trait? *J. abnorm. soc. Psychol.,* 1960, **60**, 1–7.

Cohen, A. R. Experimental effects of ego-defense preference on interpersonal relations. *J. abnorm. soc. Psychol.,* 1956, **52**, 19–27.

Crockett, W. H., & Meidinger, T. Authoritarianism and interpersonal perception. *J. abnorm. soc. Psychol.,* 1956, **53**, 378–380.

Cronbach, L. J. Processes affecting scores on "understanding others" and "assumed similarity." *Psychol. Bull.,* 1955, **52**, 177–193.

Cronbach, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford Univer. Press, 1958. Ch. 23.

Crow, W. J., & Hammond, K. R. The generality of accuracy and response sets in interpersonal perception. *J. abnorm. soc. Psychol.,* 1957, **54**, 384–396.

DeSoto, C., Kuethe, J. L., & Wunderlich, R. Social perception and self-perception of high and low authoritarians. *J. soc. Psychol.,* 1960, **52**, 149–155.

Edwards, A. L. Social desirability and the description of others. *J. abnorm. soc. Psychol.,* 1959, **59**, 434–436.

Eriksen, C. W. Perception and personality. In J. Wepman & R. Heine (Eds.), *Concepts of personality.* Chicago, Ill.: Aldine, 1963. Ch. 2.

Exline, R. V. Explorations in the process of person perception: Visual interaction in relation to competition, sex, and need for affiliation. *J. Pers.,* 1963, **31**, 1–20.

Fenichel, O. *Psychoanalytic theory of neurosis.* New York: Norton, 1945.

Feshbach, S., Singer, R. D., & Feshbach, Norma. Effects of anger arousal and similarity upon the attribution of hostility to pictorial stimuli. *J. consult. Psychol.,* 1963, **27**, 248–252.

Fiedler, F. E. Interpersonal perceptions and group effectiveness. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford Univer. Press, 1958. Ch. 16.

Foa, U. G. Convergences in the analysis of the structure of interpersonal behavior. *Psychol. Rev.,* 1961, **68**, 341–353.

Forrest, D. W., & Lee, S. G. Mechanisms of defense and readiness in perception and recall. *Psychol. Monogr.,* 1962, **76**(4, Whole No. 523).

Freud, S. The dynamics of transference. In, *Collected papers.* Vol. 2. London: Hogarth, 1924.

Gardner, R. W., Holzman, P. S., Klein, G. S., Linton, Harriet B., & Spence, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychol. Issues,* 1959, **1**(4), 1–186. (Reprinted: G. S. Klein (Ed.), *Psychological issues.* New York: Basic Books, 1959.)

Gardner, R. W., & Schoen, R. A. Differentiation and abstraction in concept formation.

Psychol. Monogr., 1962, 76(41, Whole No. 560).

GOLLIN, E. S. Organizational characteristics of social judgment: A developmental investigation. *J. Pers.*, 1958, 26, 139–154.

GROSS, CECILY F. Intrajudge consistency in ratings of heterogeneous persons. *J. abnorm. soc. Psychol.*, 1961, 62, 605–610.

HARVEY, O. J., HUNT, D. E., & SCHRODER, H. M. *Conceptual systems and personality organization.* New York: Wiley, 1961.

HASTORF, A. H. The relationships among children's categories of interpersonal perception. Paper presented at American Psychological Association, St. Louis, August 1962.

HASTORF, A. H., RICHARDSON, S. A., & DORNBUSCH, S. M. The problem of relevance in the study of person perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford Univer. Press, 1958. Ch. 5.

HATFIELD, J. A study of self-concept configurations in relation to forms of ego functioning. Unpublished doctoral dissertation, University of California, 1958.

HOLZMAN, P. S., & GARDNER, R. W. Leveling and repression. *J. abnorm. soc. Psychol.*, 1959, 59, 151–155.

ICHHEISER, G. Misunderstandings in human relations: A study in false social perception. *Amer. J. Sociol.*, 1949, 55(2, Pt. 2).

JACKSON, D. N. The measurement of perceived personality trait relationships. In N. Washburn (Ed.), *Values, decisions, and groups.* Vol. 2. New York: Pergamon Press, 1962.

JACKSON, D. N., & MESSICK, S. Individual differences in social perception. *Brit. J. soc. clin. Psychol.*, 1963, 2, 1–10.

JONES, E. E. Authoritarianism as a determinant of first impression formation. *J. Pers.*, 1954, 23, 107–127.

JONES, E. E., & DAUGHERTY, B. N. Political orientation and the perceptual effects of an anticipated interaction. *J. abnorm. soc. Psychol.*, 1959, 59, 340–349.

JONES, E. E., GERGEN, K. J., & JONES, R. G. Tactics of ingratiation among leaders and subordinates in a status hierarchy. *Psychol. Monogr.*, 1963, 77(3, Whole No. 566).

JONES, E. E., & THIBAUT, J. W. Interaction goals as basis of inference. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford Univer. Press, 1958. Ch. 11.

KATES, S. L. First impression formation and authoritarianism. *Hum. Relat.*, 1959, 12, 277–286.

KELLY, G. A. *The psychology of personal constructs.* New York: Norton, 1955. 2 vols.

KLEIN, G. S. Need and regulation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1954.* Lincoln: Univer. Nebraska Press, 1954. Pp. 224–274.

KOLTUV, BARBARA B. Some characteristics of intrajudge trait intercorrelations. *Psychol. Monogr.*, 1962, 76(33, Whole No. 552).

KOLTUV, M. Some social perceptual correlates of the valence of social behavior. Unpublished doctoral dissertation, Columbia University, 1960.

LEARY, T. *Interpersonal diagnosis of personality.* New York: Ronald Press, 1957.

LEVY, L. H. Adaptation, anchoring, and dissipation in social perception. *J. Pers.*, 1961, 29, 94–104. (a)

LEVY, L. H. The conditioning and generalization of changes in social perceptual dispositions. *J. abnorm. soc. Psychol.*, 1961, 63, 583–587. (b)

LEVY, L. H. Reversal response sets and the nature of person perception. *J. abnorm. soc. Psychol.*, 1963, 67, 392–396.

LEVY, L. H., & DUGAN, R. A constant error approach to the study of dimensions of social perception. *J. abnorm. soc. Psychol.*, 1960, 61, 21–24.

LORR, M., & MCNAIR, D. M. An interpersonal behavior circle. *J. abnorm. soc. Psychol.*, 1963, 67, 68–75.

MCDONALD, R. L. Ego control patterns and attribution of hostility to self and others. *J. Pers. soc. Psychol.*, 1965, in press.

MATKOM, A. J. Impression formation as a function of adjustment. *Psychol. Monogr.*, 1963, 77(5, Whole No. 568).

MESSICK, S., & FRITZKY, F. J. Dimensions of analytic attitude in cognition and personality. *J. Pers.*, 1963, 31, 346–370.

NABOISEK, H. Validation of a method for predicting role expectations in group therapy. Unpublished doctoral dissertation, University of California, 1953.

OMWAKE, KATHERINE T. The relation between acceptance of self and acceptance of others shown by three personality inventories. *J. consult. Psychol.*, 1954, 18, 443–446.

ORMONT, L. Tendency to differentiate in perceiving others as related to anxiety and adjustment. Unpublished doctoral dissertation, Columbia University, 1960.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PHILLIPS, E. L. Attitudes toward self and others: A brief questionnaire report. *J. consult. Psychol.*, 1951, 15, 79–81.

RABIN, H. M. Perception of others by adjusted and maladjusted subjects as reflected

in measures of perceptual space. *J. soc. Psychol.*, 1962, **56**, 149–158.

ROGERS, C. R. *Client-centered therapy*. Boston, Mass.: Houghton Mifflin, 1951.

SARBIN, T. R. Role theory. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 1. Cambridge, Mass.: Addison-Wesley, 1954. Ch. 6.

SCHULBERG, H. C. Authoritarianism, tendency to agree, and interpersonal perception. *J. abnorm. soc. Psychol.*, 1961, **63**, 101–108.

SCODEL, A., & FRIEDMAN, MARIA L. Additional observations on the social perception of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1956, **52**, 92–95.

SCODEL, A., & MUSSEN, P. Social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1953, **48**, 181–184.

SECHREST, L., & JACKSON, D. N. Social intelligence and accuracy of interpersonal predictions. *J. Pers.*, 1961, **29**, 167–182.

SECORD, P. F. Facial features and inference processes in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. Ch. 20.

SECORD, P. F., BACKMAN, C. W., & MEREDITH, HELEN E. Cue-dominance in person perception as a function of strength of perceiver-need. *J. soc. Psychol.*, 1962, **58**, 305–313.

SECORD, P. F., & BERSCHEID, ELLEN S. Stereotyping and the generality of implicit personality theory. *J. Pers.*, 1963, **31**, 65–78.

SECORD, P. F., DUKES, W. F., & BEVAN, W. Personalities in faces: I. An experiment in social perceiving. *Genet. Psychol. Monogr.*, 1954, **49**, 231–279.

SECORD, P. F., & JOURARD, S. M. Mother concepts and judgments of young women's faces. *J. abnorm. soc. Psychol.*, 1956, **52**, 246–250.

SECORD, P. F., & MUTHARD, J. E. Individual differences in the perception of women's faces. *J. abnorm. soc. Psychol.*, 1955, **50**, 238–242.

SINGER, R., & FESHBACH, S. Effects of anxiety arousal in psychotics and normals upon the perception of anxiety in others. *J. Pers.*, 1962, **30**, 574–587.

SMELSER, W. T. Dominance as a factor in achievement and perception in cooperative problem solving interactions. *J. abnorm. soc. Psychol.*, 1961, **62**, 535–542.

STEINER, I. D., & JOHNSON, H. H. Authoritarianism and "tolerance of trait inconsistency." *J. abnorm. soc. Psychol.*, 1963, **67**, 388–391.

SUINN, R. M. The relationship between self-acceptance and acceptance of others: A learning theory analysis. *J. abnorm. soc. Psychol.*, 1961, **63**, 37–42.

SULLIVAN, H. S. *Conceptions of modern psychiatry*. Washington, D. C.: William Alanson White Psychiatric Foundation, 1948.

TAFT, R. The ability to judge people. *Psychol. Bull.*, 1955, **52**, 1–23.

TAGIURI, R. Introduction. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. (a)

TAGIURI, R. Social preference and its perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. Ch. 21. (b)

THIBAUT, J. W., & RIECKEN, H. W. Authoritarianism, status, and the communication of aggression. *Hum. Relat.*, 1955, **8**, 95–120.

TRIPODI, T., & BIERI, J. Cognitive complexity as a function of own and provided constructs. *Psychol. Rep.*, 1963, **13**, 26.

TUCKER, L. R. Implications of factor analysis of three-way matrices for measurement of change. In C. W. Harris (Ed.), *Problems in measuring change*. Madison: Univer. Wisconsin Press, 1963.

WILKINS, E. J., & DECHARMS, R. Authoritarianism and response to power cues. *J. Pers.*, 1962, **30**, 439–458.

WITKIN, H. A., DYK, RUTH B., FATERSON, HANNA F., GOODENOUGH, D. R., & KARP, S. A. *Psychological differentiation*. New York: Wiley, 1962.

WOLIN, L. R. An analysis of the content of interpersonal perception. Unpublished doctoral dissertation, Cornell University, 1956.

(Received February 17, 1964)

Psychological Bulletin
1964, Vol. 62, No. 5, 309-322

THE INTERRUPTION OF TASKS:

METHODOLOGICAL, FACTUAL, AND THEORETICAL ISSUES [1]

EARL C. BUTTERFIELD

Yale University

After describing the criterion scores which have been used to assess behavior in the interrupted task paradigm (ITP), a summary of the research literature is presented. ITP as a source of data for evaluating the psychoanalytic theory of repression is found not to allow for the separation of learning and retention effects, and so is not well suited to the study of repression. Similarly, the mediation-avoidance hypothesis makes predictions only concerning interrupted task recall and while it is partially consistent with indirect data, it has yet to receive direct experimental test. The need-achievement conception predicts interrupted task recall satisfactorily but is inapplicable to completed task recall or relative recall scores. Finally ITP is considered as a source of data for a developmental conception of success-failure reactions. Repetition choice scores are found to be consistent with the developmental theory, but some recall results are not.

In the interrupted task paradigm (ITP) the subject (*S*) is engaged in a number of tasks some of which he is allowed to complete and some of which are interrupted before he completes them. Zeigarnik (1927) first introduced ITP as a test of the prediction from the gestalt theory of motivation that interrupted tasks should be recalled more frequently than completed tasks. Her findings supported this prediction and is known to this day as the Zeigarnik effect. However, ITP encompasses more than just the Zeigarnik effect or even the recall of tasks. In the first place, the Zeigarnik effect is far from being the invariable result in ITP. Frequently, more completed than incompleted tasks are recalled (e.g., Atkinson, 1953). In the second place, measuring the relative recall of completed and incompleted activities is not the only way to assess behavior in ITP. For example, ITP may be assessed by having *S* choose which task he would like to repeat (e.g., Rosenzweig, 1933, 1945). Finally, and most important, ITP has assumed a much wider theoretical significance than just being a test of a derivation from gestalt theory. (The reader who is interested in ITP as a source of evidence for the differential utility of gestalt and stimulus-response theory should see Osgood, 1951.) It has been used to test the psychoanalytic theory of repression (Rosenzweig, 1938), a mediation-avoidance hypothesis of personality functioning (Inglis, 1961), the achievement-motive conceptions of McClelland, Atkinson, Clark, and Lowell (1953), and a developmental theory of success-failure conceptualization (Cromwell, 1963). Indeed, ITP has become one of those instances in the history of psychology when a single technique has been used to test several theoretical issues. Therefore, the methodological and theoretical issues involved and the

[1] This paper is a revision of part of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Peabody College. At the time of this revision the author was supported by Research Grant MH-06809 from the National Institute of Mental Health, United States Public Health Service.

factual findings with ITP deserve elucidation.

Because there has been a great deal of inconsistency and circularity in the interpretation of findings with ITP (Alper, 1952; Glixman, 1948; Osgood, 1951), this review first summarizes the functional relationships found with ITP and then considers the theoretical issues which have grown out of these findings.

ITP Criterion Scores

The empirical findings from ITP can most readily be grasped by grouping them according to criterion scores. It will subsequently be seen that such a grouping also clarifies the theoretical issues. The criterion scores used in ITP studies fall naturally into three groups. This grouping excludes studies which have used as their criterion scores either the relative rate of spontaneous resumption of incompleted and completed tasks (Henle, 1942; Nowlis, 1941; Rickers-Ovsiankina, 1937) or attractiveness ratings (Cartwright, 1942; Gebhardt, 1948) because there are so few of them and because the factors which determine spontaneous resumption seem so numerous and so poorly understood.

One group of criterion scores which is concerned with recall of ITP tasks keeps the recall of incompleted (I) and completed (C) tasks separate. Thus, the measures are the number of I or C tasks recalled (IR and CR, respectively).

A second group of ITP criterion scores which is also concerned with task recall includes only those measures which reflect the relative recall of I and C tasks taken together. This group includes the various recall differences and recall ratio scores. The recall difference scores are: IR-CR, CR-IR, and CR/C-IR/I. The recall ratio scores are: IR/CR, $\frac{IR/I}{CR/C}$, and a variety of more complex and less interpretable ratios.

The feature common to all relative recall scores is that changes in them may not be attributed to any particular change in either IR or CR individually (Glixman, 1948). An increase in a difference or ratio score may be due to: an increase in IR accompanied by no change in CR, a decrease in CR accompanied by no change in IR, an increase in IR accompanied by a decrease in CR, an increase in IR accompanied by a smaller increase in CR, and a decrease in CR accompanied by a smaller decrease in IR. Likewise, the opposite of these five conditions can lead to a decrease in relative recall score. Therefore, either an increase or a decrease in a recall ratio or difference score may reflect an increase, no change, or a decrease in IR or CR individually. Furthermore, a relative recall score may remain constant even when both IR and CR change. For these reasons, hypotheses concerned with change in either IR or CR alone must measure them directly rather than by relative recall scores. Also, any theoretical scheme which makes predictions about only IR or CR cannot be tested by a study which employs only relative recall scores. Only when a theory makes predictions about relative differences or ratios, or when it predicts both IR and CR separately, is a relative recall score appropriate to it. Even then the analysis of preference is one which does not use a relative recall score. The preferred procedure is to use completion-incompletion as a dimension of the analysis. When completion-incompletion is used as a dimension of the analysis no information about either CR or IR is lost. Consequently, any relative recall effects, which are mani-

fested as interactions involving the completion-incompletion dimension, are immediately interpretable. That is, CR and IR are not obscured in this procedure as they are in the calculation of relative recall scores.

A third group of ITP criterion scores is concerned with task repetition rather than task recall. Task repetition scores are secured by requiring *S* to choose to repeat one member of a pair of tasks, one of which he has previously completed and one of which he has previously begun but not completed. In all but two investigations of repetition choice (RC) to date (Butterfield, 1963; Stedman, 1962), *S*s have been given only two tasks, one of which was completed and one of which was interrupted. Consequently, the only score used to measure RC has been some indication of whether *S* repeated the I or the C task. Therefore, no relative recall scores, such as ratios and differences, have been employed.

Investigators have typically used either recall or repetition-choice scores without any consideration of the relationships between the two classes of criterion scores. This fact reflects an astonishing lack of basic methodological work with ITP. Despite the long history of research with ITP, Butterfield (1963) was apparently the first to examine the intercorrelations of the various criterion scores which it yields. Yet, knowledge of this kind is essential if the literature is to be integrated. More crucially, there has apparently been no systematic investigation of the reliability of the various ITP scores.

Correlates of the Various ITP Criterion Scores

IR and CR Taken Separately

A substantial portion of the experimentation with ITP has concerned the effects of instructions administered prior to task presentation. In practically all of these investigatons the instructions have varied along a dimension which Alper (1946) characterized as task versus ego orientation and which Inglis (1961) characterized as stressful versus nonstressful. Under the ego or stress instruction, *S* is told that task completion or incompletion is a measure of his ability or intelligence. Under the task-oriented or nonstress instruction, *S* is told that task completion or incompletion is of secondary importance and that it has nothing to do with his ability or intelligence. In the present article this instructional dimension is referred to as neither an ego versus task-orientation nor a stressful versus nonstressful continuum. It is referred to as a skill versus nonskill dimension because this latter label emphasizes the nature of the instructions rather than *S*'s hypothetical reaction to the instructions.

The most striking thing about the various findings on the effects of skill instructions is their inconsistency. Skill instructions have been shown to increase (Atkinson, 1953), to have no effect upon (Alper, 1957; Atkinson & Raphelson, 1956; Eriksen, 1952a; Kendler, 1949; Rosenzweig, 1943), and to decrease IR (Alper, 1946; Caron & Wallach, 1957; Eriksen, 1952b, 1954; Glixman, 1949; Smock, 1956). Similarly with skill instructions, CR has been shown to decrease (Alper, 1946; Caron & Wallach, 1957; Kendler, 1949), to remain the same (Alper, 1957; Atkinson, 1953, Atkinson & Raphelson, 1956; Eriksen, 1952b, 1954; Glixman, 1949; Rosenzweig, 1943; Smock, 1956), and to increase (Eriksen, 1952a).

Several studies suggest that the inconsistent findings concerning the effects of skill instructions upon IR and CR, particularly upon IR, may be due to differences in *S* variables between the

studies. Atkinson (1953) found that high need achievement (n Ach) *S*s increased their recall of incompleted tasks while low n Ach *S*s decreased their recall of incompleted tasks when instructions were more skill oriented. Atkinson and Raphelson (1956) found no differences between high and low n Ach *S*s under nonskill instructions. However, under skill instructions high n Ach *S*s recalled more I tasks than low n Ach *S*s. Atkinson and Raphelson also found that while there were no differences in high and low need affiliation *S*s under skill instructions, high need affiliation *S*s recalled more I tasks under nonskill instructions than did low need affiliation *S*s. In these nonskill instructions, *S* was told that he was helping the experimenter participating in ITP. Neither Atkinson nor Atkinson and Raphelson found any instruction—or personality—related differences in the recall of completed tasks. Caron and Wallach (1959) replicated Atkinson and Raphelson's findings concerning IR and n Ach, even though they used a substantially different measure of n Ach. In addition, however, Caron and Wallach found that while high n Ach *S*s' recall of completed tasks was unaffected by skill instructions, low n Ach *S*s recalled fewer completed tasks under skill than nonskill instructions. Eriksen (1954) found that *S*s who were high on the MMPI psychasthenia recalled more I tasks under skill instructions than *S*s who were low on that scale, and that *S*s who were high on the MMPI hysteria scale recalled fewer I tasks under skill instructions than *S*s low on that scale. Eriksen found no personality-related differences in the recall of C tasks. Eriksen also found that *S*s with ego strength, as inferred from high $F+\%$ on the Rorschach, recalled more I tasks under skill instructions than low ego-strength *S*s. Using a different Rorschach measure of ego strength, Jourard (1954) found no relationship with IR or CR. Apparently high n Ach and strong-ego *S*s respond to skill instructions by recalling more I tasks while n Ach and weak-ego *S*s respond to skill instructions by recalling fewer I tasks. The trend for CR is less clear. The degree of overlap in the n Ach and ego-strength constructs remains in question.

Marrow (1938) investigated instructions which did not vary along the typical skill versus nonskill dimension. Some *S*s were told task completion indicated success; others were told task incompletion indicated failure. His results clearly indicated that *S*'s interpretation of C and I, rather than C and I per se, determined recall. Findings by McKinney (1935) support this interpretation. Subsequent analysis of Mardow's data also indicated that IR and CR were both unaffected by telling *S* during the course of ITP that he was doing very poorly on ITP as a whole.

The relationships of IR and CR to variables other than instructional differences have also been investigated. For example, IR and CR are differently affected by the passage of time. Martin (1940) found an increase in the recall of C tasks 2 days after task presentation which was not present for I tasks. Pachuri (1935) found a similar increase 1 day after task presentation. After 2 weeks, however, Martin (1940) found that the reminiscence of C tasks had dissipated and that both IR and CR had decreased. Sanford and Risser (1948) found after both 2 weeks and 4 months that CR had decreased significantly while IR had not changed.

Pachuri (1935) found that fatigue at the time of task administration reduced IR but did not affect CR and that fatigue at the time of recall affected neither IR nor CR.

Caron and Wallach (1957) found that the verbal cancellation or relief of skill instructions subsequent to task recall did not affect IR. Subjects who had and *S*s who had not had their skill instructions relieved recalled the same number of I tasks, this being fewer than that of nonskill-instruction *S*s.

Recall of I Relative to C Tasks

The experimental findings concerning the relative recall of I and C tasks can best be summarized under the following four headings: effects of instructions, effects of *S* variables, interaction of instructions and *S* variables, and other variables.

Effects of Instructions. The experimental literature unequivocally indicates that skill instructions increase the ratio of C task recalled to I task recall (Alper, 1946; Eriksen, 1952a, 1952b; Hays, 1952; Kendler, 1949; Lewis, 1944; Lewis & Franklin, 1944; Rosenzweig, 1941, 1943; Smock, 1956).

Effects of S *Variables.* Rosenzweig and Sarason (1942) found that *S*s who were more suggestible and/or hypnotizable recalled relatively more C than I tasks while the opposite was true of *S*s who were less suggestible and hypnotizable. In a related study Sarason and Rosenzweig (1942) found, by means of ratings derived from Thematic Apperception Test protocols, that those *S*s who recalled more C than I tasks were relatively high on need affiliation and need deference while those *S*s who recalled more I than C tasks were relatively high on need autonomy and on anxiety.

Tamkin (1957) found that under skill instructions more adult schizophrenics than adult normals recalled more I than C tasks. Winder (1952) found that under skill instructions more paranoid schizophrenics than nonparanoid schizophrenics recalled more I than C tasks.

Sanford (1946) and Rosenzweig and Mason (1934) have investigated the effects of mental age (MA) and chronological age (CA) upon relative recall. Using skill instructions, crippled institutionalized *S*s, and different numbers of tasks for different *S*s, Rosenzweig and Mason failed to find any clear relationship between either MA or CA and relative recall. Nevertheless, they reported an impression that as both MA and CA increased the recall of C tasks increased to the recall of I tasks. Stanford, on the other hand, used nonskill instructions with slightly older, intellectually and physically average children from the public schools. He found a clear increase in the recall of I relative to C tasks as a function of CA. Stanford's findings are probably more general and reliable than Rosenzweig and Mason's. Butterfield's (1963) findings are consistent with this conclusion.

Interaction of Instruction and S *Variables.* Instructions and *S* variables interact when relative recall scores are used as the criterion. Persons with strong egos as measured by a composite Rorschach score (Eriksen, 1954) and a questionnaire (Alper, 1957) recall relatively more I tasks under nonskill instructions and relatively more C tasks under skill instructions. Alper's strong-ego *S*s have higher need recognition and need dominance and react to failure with more increased effort than do her weak-ego *S*s (Alper, 1948). People with low n Ach recall relatively more I tasks under nonskill instructions and relatively more C tasks under skill instructions (Atkinson, 1953; Atkinson & Raphelson, 1956; Caron & Wallach 1957, 1959). People with weak egos and high n Ach recall relatively more C tasks under nonskill instructions and

relatively more I tasks under skill instructions.

Other Variables. Abel (1938) investigated the relationship of relative recall to the Schneider index of neurocirculatory efficiency. The Schneider index is supposedly a measure of response to physiological stress. Abel found that *S*s who recalled more I tasks relative to C tasks had significantly greater neurocirculatory efficiency. Strother and Cook (1953) have shown that *S*s with low neurocirculatory efficiency respond to stress with a decrement in performance. Therefore, it may be that *S*s who recall more I than C tasks respond with less decrement in performance under stress than do *S*s who recall more C than I tasks. Bialer and Cromwell (1960) found evidence which seems to support the suggestion that *S*s who recall more I than C tasks respond with less performance decrement to stress than *S*s who recall more C than I tasks. They found that mentally retarded *S*s who return to incompleted tasks respond with a greater increase in rate of card sorting after induced failure than do *S*s who return to completed tasks in ITP procedure.

Hays (1952) found that the interpolation of an interesting task between the ITP and recall caused *S*s to recall relatively fewer C tasks than when he interpolated a boring task.

RC

The RC has been studied exclusively with children and has been most extensively related to MA and CA. Bialer (1957), Bialer and Cromwell (1960), Butterfield (1963), Rosenzweig (1933, 1945), and Spradlin (1955–60) have shown that as both MA and CA increase more *S*s choose to repeat I tasks. Crandall and Rabson (1960) have made similar findings for boys but not for girls. Bialer (1960) has found by using average and mentally retarded *S*s that this increasing tendency to choose I instead of C tasks under skill instructions is due solely to increase in MA. It is unrelated to increases in CA when MA is partialed out. Miller (1961) found that RC of the I task varies with MA and not with CA when ITP is administered under skill instructions but not under nonskill instructions. Bialer (1957, 1960), Bialer and Cromwell (1960), Butterfield (1963), Rosenzweig (1933, 1945), Spradlin (1955–66), and Crandall and Rabson (1960) all used skill instructions so that Miller's (1961) findings are consistent with theirs. Miller's failure to find a relationship between RC and MA under nonskill conditions may be due to the fact that practically all of his *S*s chose the I task under nonskill conditions. Consequently, there was no possibility of his finding any differential relationship between the choice of I and C tasks.

A number of apparently similar personality variables have been related to RC behavior. Subjects who return to I rather than C tasks have been rated by teachers as more rebellious (Miller, 1961) and as having more pride (Rosenzweig, 1933) and by observers of their free-play activity as being more assertive and independent (Crandall & Rabson, 1960). Bialer (1960) found that *S*s who reported in questionnaire items that they feel personal control over what happens to them more often repeat I tasks than *S*s who report feelings of more external control. It is possible that none of these relationships would have been found if MA had been controlled. For example, Bialer's (1960) finding was eliminated when MA was controlled statistically.

In the only study to investigate the effects of different instructions upon RC, Miller (1961) found that skill

instructions increased the tendency to return to C rather than I tasks.

Theoretical Issues

The ITP has been considered to provide validity measures for the psychoanalytic theory of repression, the mediation-avoidance hypothesis of personality functioning, the McClelland and his associates' (1953) conception of achievement motivation, and a developmental theory of success-failure conceptualization. Its role in each of these four theoretical realms is considered in turn.

Repression

Rosenzweig (1938) originated the rationale for deriving a measure of repression from ITP. He reasoned that skill instructions would make I tasks seem like failure to the *S*s, thereby causing them to repress the I tasks. Therefore, *S*s would recall fewer I tasks under skill instructions than under nonskill instructions. Rosenszweig derived no predictions from psychoanalytic theory about the effects of skill instructions upon C tasks and no such derivation seems possible.

Rosenzweig found that the ratio of I tasks to C tasks was less under skill than under nonskill conditions. However, Rosenzweig suggested, and Glixman (1948) subsequently established by reanalyzing Rosenzweig's data, that this relative change was due to an increase in CR rather than to the predicted decrease in IR. Other investigators have also failed to find any consistent effects of skilled instructions upon IR when *S*s were unselected for personality variables (Alper, 1946, 1957; Atkinson, 1953; Atkinson & Raphelson, 1956; Caron & Wallach, 1957; Eriksen, 1952a, 1952b, 1954; Glixman, 1949; Kendler, 1949). When *S*s were selected for personality variables, skill instructions increased the IR of *S*s high in n Ach and MMPI psychasthenia and decreased the IR of *S*s low in n Ach and high in MMPI hysteria. Therefore, if a decrease in IR due to skill instructions is indeed repression, repression occurs markedly more often in some personalities than in others. This is as might be expected. Furthermore, other personalities actually show the opposite of repression—call it obsession (Caron & Wallach, 1959)—to task interruption under skill conditions.

While it is conceivable that some *S*s should react to interruption under skill instructions by repressing the I task and others by becoming obsessed with the I task, there is a good reason to doubt that changes in IR due to skill instructions have demonstrated either repression or obsession. While the present argument is made solely for the phenomenon of repression, it also applies to obsession. It is said to occur when some experience has been registered, that is, learned but because of some painful association has been warded off from recall. In order to demonstrate repression, it is first necessary to demonstrate that two groups have learned some response equally well and then to show that some painful experience leads one group to recall that response significantly less well. That is, it is necessary to be able to separate the effects of learning from the effects of retention upon recall. In ITP, it is impossible to show that the groups to be compared on retention are equal on original learning. For an example of a study which separates the effects of learning and retention by placing the repression-inducing stimulus after a test for initial learning see Zeller (1952). This inability to show that I and C tasks are equally well learned is a crucial shortcoming of ITP as a measure of retention since the original learning opportunity is frequently shorter

for I than for C tasks. Since ITP does not allow for the separation of learning from retention it is not well suited to the study of repression. Nevertheless, indirect attempts have been made to determine whether differential recall of I tasks was a phenomenon of learning or of retention.

Theoretically, repressed material should not be subject to the same laws of retention as unrepressed material as long as the reason for the repression exists. In ITP, the ostensible reason for repression is *S*'s reaction to his belief that task incompletion is due to his lack of skill. Therefore, until the skill instructions are somehow countermanded and *S* "realizes" that task incompletion is unrelated to his skillfulness, IR should not change over time the way CR does. However, there should be no differences in changes over time between IR and CR when the tasks are presented under nonskill instructions. As was reported earlier, Martin (1940), Pachuri (1935), and Sanford and Risser (1946) found differences between changes in IR and changes in CR, but these differences occurred under both skill and nonskill instructions. Therefore, these differences establish neither that decreases in IR are due to retention, that is, repression, nor that they are due to original learning.

Caron and Wallach (1959) theorized that, since the ostensible reason for repression in ITP is *S*'s reaction to his belief that task incompletion is due to his lack of skill, they could determine whether differential recall was due to learning or retention by comparing the recall of three groups: nonskill instructions, skill instructions, and skill instructions which were countermanded between the time of task presentation and task recall. If the countermanded skill-instruction group was similar to the nonskill group, then the effect would be one of retention. If the countermanded skill group was similar to the skill group and different from the nonskill group, then the effect would be one of original learning. Their results suggested that differences were probably due to differences in original learning.

Pachuri's (1935) finding that fatigue at the time of task administration affects IR while fatigue at the time of task recall does not, also suggests that depressed recall is due to differences in learning rather than differences in retention.

In summary, both logical consideration and research findings with ITP suggest that, as it has been used in the past, it is not well suited to the study of repression.

Mediation-Avoidance Hypothesis

Inglis (1961) attempted to integrate the results of ITP studies by means of the mediation-avoidance hypothesis. According to this hypothesis, *S*s react to moderate amounts of "socialized anxiety" by avoiding the mediator or instigator of that anxiety. Small amounts of anxiety are said to have no influence upon avoidance behavior and large amounts are said to be disruptive of avoidance behavior. In other words, the mediation-avoidance hypothesis postulates a curvilinear relationship between anxiety and avoidance. Furthermore, there are said to be two sources of socialized anxiety: *S*'s immediate environment and his personality. The environment ranges from nonstressful to stressful situations in such a way that task interruption under skill instructions is more stressful than task interruption under nonskill instructions. The personality factors which are said to be related to socialized anxiety are the dimensions of stability to neuroticism and introversion to extraversion.

Neurotic extraverts are least susceptible, neurotic introverts are most susceptible, and stable introverts and stable extraverts are intermediately susceptible to stress. Since, according to the mediation-avoidance hypothesis, there is a curvilinear relationship between anxiety and avoidance, and since anxiety is contributed to by both the environment and *S*'s personality, the mediation-avoidance hypothesis predicts an interaction in the production of avoidance reactions between personality and situational stress (see Figure 1).

From these premises of the mediation-avoidance theory the present author has derived the following predictions.

1. Under conditions of low stress, introverts recall fewer I tasks than extraverts.

2. Under conditions of high stress, introverts recall more I tasks than extraverts.

3. If both the introverts and extraverts are neurotic the differences predicted in Hypotheses 1 and 2 above are greater than if the groups are unselected for neuroticism.

4. Under similar instructional conditions, there are no mean differences between neurotic and stable groups which are equal on or unselected for introversion-extraversion.

5. Under similar instructional conditions, the variance is greater and the distribution of avoidance scores is more bimodal for neurotic than for stable groups when the groups are equal on extraversion-introversion.

6. The effects of stress instructions upon IR are unpredictable without knowledge of *S*'s personality.

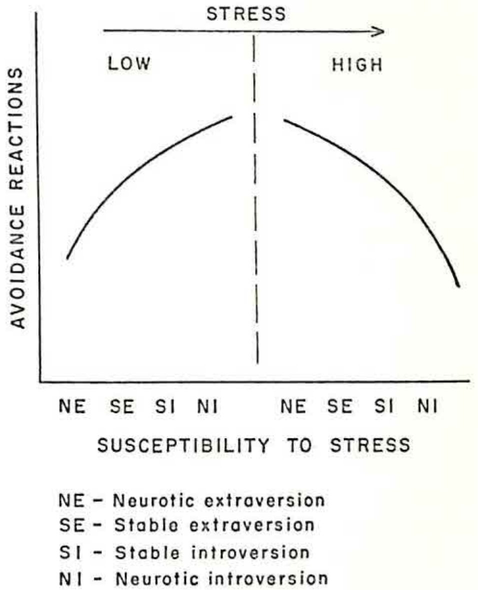


FIG. 1. Hypothetical relationship between susceptibility to stress and avoidance reactions (adapted from Inglis, 1961).

These predictions refer only to the recall of interrupted tasks because according to Inglis (1961, p. 280) I tasks are the only "anxiety-producing mediators" in ITP. Therefore, since changes in relative recall scores depend upon changes in both IR and CR, relative recall scores do not provide a definitive test of the validity of the mediation-avoidance hypothesis. Consequently, Inglis' (1961) discussion of ITP in which he relied entirely upon relative recall scores is inconclusive with regard to the mediation-avoidance hypothesis and his conclusion that ITP studies confirm the mediation-avoidance hypothesis is questionable if not erroneous.

There are no studies in the literature which test directly the predictions made from the mediation-avoidance hypothesis. However, some indirect tests have been made. For example, if it is assumed, as Inglis does, that high n Ach is an index of neurotic extraversion, then the findings of Atkinson (1953), Atkinson and Raphelson (1956), and Caron and Wallach (1959) support Hypotheses 1 and 2 above. Since Hypotheses 1 and 2 taken together constitute the prediction of an

interaction between instructional and S variables, the mediation-avoidance hypothesis predicts some of the most complex ITP phenomena. Hypothesis 3 has not been tested even indirectly. If it is assumed, as Inglis does, that Eriksen's (1954) measure of ego strength is a valid index of neuroticism, then Hypothesis 4 has been contradicted and Hypothesis 5 has received no support. The author's conclusion that Hypothesis 4 has been contradicted differs from Inglis' interpretation of the evidence. Inglis concluded that a study by Eriksen (1954) supported the mediation-avoidance hypothesis when it would not have done so even if a relative recall score was an appropriate criterion score with which to test the hypothesis. Inglis noted that Eriksen measured ego strength "partly by a Rorschach conformity score not unlike that found by Eysenck to be a fairly good measure of neuroticism." He concluded from the fact that Eriksen's ego-strength measure was positively related to relative recall scores under nonstress conditions and negatively related to relative recall scores under stress conditions that the mediation-avoidance hypothesis was upheld. It is clear from Inglis' diagrams (see Figure 1), however, that neuroticism independent of extraversion-introversion is predicted to be unrelated to directional differences in avoidance. That is, Eriksen's finding of a directional relationship between neuroticism and relative recall is contrary to rather than supportive of the derivation from the mediation-avoidance hypothesis. Hypothesis 6 has been supported by the lack of consistency in the results of studies examining the effects of skill instructions upon IR.

It is apparent that while the mediation-avoidance hypothesis is relevant to the recall of I tasks, much more research is needed before it can be asserted that it adequately summarizes or predicts even the existing findings with ITP.

Achievement Motivation

In order to predict ITP findings from n Ach scores it has been assumed that differences in n Ach indicate qualitative differences in motive structure rather than, or in addition to, differences in intensity of motivation (McClelland, 1951; McClelland et al., 1953). Specifically, it has been assumed that persons with high n Ach scores are motivated primarily by a need to approach succcess while persons with moderate or low n Ach scores are motivated primarily by a fear of, or a need to, avoid failure. Furthermore, skill instructions are assumed to enhance the motivations which the person possesses. That is, for low or moderate need achievers, skill instructions are said to be threatening and to lead to increased avoidance of failures, that is, I tasks "since recall of failures would serve to redintegrate the pain of failure [McClelland et al., 1953]." On the other hand, for high need achievers, skill instructions are said to be an incentive or challenge which leads to increased approach to, or recall of, I tasks. It is as if the high need achiever "wanted to continue to strive to complete them [McClelland, et al., 1953]." These predictions about the interactive effects of instructions and n Ach on IR are supported by the experimental literature.

As it was developed by McClelland and his associates (1951, 1953), the achievement-motive conception makes no predictions about the recall of C tasks. Therefore, it cannot predict the interactive effects of instructions and n Ach upon relative recall scores without the addition of factually unwarranted assumptions. It would be neces-

sary to assume that neither n Ach nor skill instructions affect CR. While it is true that n Ach is unrelated to CR, instructions do affect it. Consequently, without the addition of some specification of conditions under which instructions affect CR, achievement-motive theory is concerned with only the relationship between IR and n Ach.

Success-Failure Conceptualization

While theorists who were interested in repression, the mediation-avoidance hypothesis, and achievement motivation have used task recall measures as their criterion scores, theorists who were interested in success and failure conceptualization and developmental variables have used task RC.

These latter theorists (Bialer, 1960; Cromwell, 1963) have based their predictions about RC on developmental hypotheses such as the following:

From the time of birth the child is assumed to respond to stimuli as pleasant or unpleasant and to learn to approach stimuli previously associated with pleasant events and to avoid stimuli previously associated with unpleasant events. As the child matures a second motivational and conceptual system is said to develop. This second system is called the success-approach versus failure-avoidance system. As it matures the child comes more and more to conceptualize events in terms of his own behavioral competence, that is, success and failure and less and less in terms of pleasantness and unpleasantness. In terms of the ITP situation in which the child is asked to repeat the task of his choosing, this means that the older child will more often choose to repeat the task which allows him to demonstrate behavioral competence. The younger child, on the other hand, will more often choose the task which gave him the greatest pleasure. Since more behavioral competence can be demonstrated by doing a more difficult task, the older child will more often choose the more difficult task. Theoretically, the only clue to difficulty in the ITP situation is completion or incompletion, the I task appearing to be more difficult. Therefore, the older child should more often return to the I task even though the C task was probably the more pleasurable to him. The evidence strongly supports this prediction.

This developmental viewpoint has difficulty predicting the relationship between ITP behavior and instructional and *S* variables. The predictions it does make all stem from a straightforward and well-documented derivation from the notion of a developmental success-striving versus failure-avoiding motivation system; namely, that as children mature they report more frequently and in a wider variety of situations that they feel control over what happens to them (Bialer, 1960; McConnell, 1963; Miller, 1961). To the extent that this tendency to report oneself in control of what happens to him is correlated with MA or CA, the increases in it are associated with increasing resumption and recall of I tasks. However, there is some evidence that the tendency to verbalize control is not solely dependent upon MA or CA and that increases in the portion of verbalized control which are independent of MA or CA are associated with increasing resumption of C tasks (Bialer, 1960). These findings have been interpreted (Butterfield, 1963) to mean that the particular part of the variance in verbalized control which is independent of MA is a measure of the degree to which an individual strives to avoid events at which he succeeded in the past. This interpretation gives the developmental position two *S* variables by which to predict ITP be-

havior: intellectual maturity and verbalized control. As intellectual maturity increases, the resumption or recall of I tasks becomes more frequent. As an individual verbalizes more control over events without concomitant growth in intellectual maturity, the resumption or recall of C tasks becomes more frequent.

The verbalized control construct also makes it possible for the developmental theory to make predictions about the effects of skill instructions. According to the developmental position, the effect of skill instructions is the same as that of *S*'s verbalized control. That is, skill instructions simply induce *S* to believe that he controls task completion. Internal control independent of MA leads to reduced resumption of I and increased resumption of C tasks.

Virtually all of the data within the realm of task resumption are consistent with the developmental theory outlined above. However, it does not predict the more thoroughly studied task recall phenomena as well. Skill instructions do not always decrease the recall of I tasks and increase the recall of C tasks as the developmental theory predicts. Furthermore, the developmental theory does not include variables with known relationships to many of the *S* variables which have been shown to interact with the various ITP scores. For example, it can make no predictions about the effects of n Ach. Therefore, it is not applicable to some of the most reliable and orderly findings with ITP. However, the developmental position does predict the fact that skill instructions always decrease the recall of I tasks to the recall of C tasks and vice versa. This is a distinct difference between it and the other theoretical schemes considered here. The reason the developmental scheme predicts the effects of skill instructions upon relative recall scores is that, unlike all of the other positions here considered, it makes predictions about both IR and CR. Consequently, relative recall scores are as relevant to the development position as are scores which keep IR and CR separate.

Conclusion

It is apparent that none of the systematic attempts to deal with ITP phenomena has adequately accounted for all of the known relationships. However, these theoretical efforts and the literature upon which they are based clearly indicate that any adequate explanation of ITP phenomena must take into account: instructions, personality variables, developmental level of the subject, and probably task variables.

REFERENCES

Abel, Theodora M. Neurocirculatory reaction and the recall of unfinished and completed tasks. *J. Psychol.*, 1938, **6**, 377–383.

Alper, Thelma G. Memory for completed and incompleted tasks as a function of personality: An analysis of group data. *J. abnorm. soc. Psychol.*, 1946, **41**, 403–420.

Alper, Thelma G. Memory for completed and incompleted tasks as a function of personality: Correlation between experimental and personality data. *J. Pers.*, 1948, **17**, 104–137.

Alper, Thelma G. The interrupted task method in studies of selective recall: A re-evaluation of some experiments. *Psychol. Rev.*, 1952, **59**, 71–88.

Alper, Thelma G. Predicting the direction of selective recall; its relation to ego strength and *N* achievement. *J. abnorm. soc. Psychol.*, 1957, **55**, 149–165.

Atkinson, J. W. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 381–390.

Atkinson, J. W., & Raphelson, A. C. Individual differences in motivation and behavior in particular situations. *J. Pers.*, 1956, **24**, 349–363.

Bialer, I. Differential task recall in mental defectives as a function of consistency in self-labeling. Unpublished master's thesis, George Peabody College for Teachers, 1957.

Bialer, I. *Conceptualization of success and failure in the mentally retarded and normal children.* Ann Arbor, Mich.: University Microfilms, 1960. (Also, in brief, *J. Pers.*, 1961, **29**, 303–320.)

Bialer, I., & Cromwell, R. L. Task repetition in mental defectives as a function of chronological and mental age. *Amer. J. ment. Defic.*, 1960, **65**, 265–268.

Butterfield, E. C. *Task recall and repetition choice as a function of locus of control components, mental age, and skill vs. nonskill instructions.* Ann Arbor, Mich.: University Microfilms, 1963.

Caron, A. J., & Wallach, M. A. Recall of interrupted tasks under stress: A phenomena of memory or learning? *J. abnorm. soc. Psychol.*, 1957, **55**, 372-381.

Caron, A. J., & Wallach, M. A. Personality determinants of repressive and obsessive reactions to failure stress. *J. abnorm. soc. Psychol.*, 1959, **59**, 236–245.

Cartwright, D. The effect of interruption, completion, and failure upon the attractiveness of activities. *J. exp. Psychol.*, 1942, **31**, 1–16.

Crandall, V., & Rabson, Alice. Children's repetition choices in an intellectual achievement situation following success and failure. *J. genet. Psychol.*, 1960, **97**, 161–168.

Cromwell, R. L. A social learning approach to mental retardation. In N. R. Ellis (Ed.), *Handbook of mental retardation.* New York: McGraw-Hill, 1963.

Eriksen, C. W. Defense against ego threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, **47**, 231–236. (a)

Eriksen, C. W. Individual differences in defensive forgetting. *J. exp. Psychol.*, 1952, **44**, 442–443. (b)

Eriksen, C. W. Psychological defenses and "ego strength" in recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45–50.

Gebhardt, M. E. The effect of success and failure upon the attractiveness of activities as a function of experience, expectation, and need. *J. exp. Psychol.*, 1948, **38**, 371–388.

Glixman, A. F. An analysis of the use of the interruption technique in experimental studies of "repression." *Psychol. Bull.*, 1948, **45**, 491–506.

Glixman, A. F. Recall of completed and incompleted activities under varying degrees of stress. *J. exp. Psychol.*, 1949, **39**, 281–295.

Hays, D. G. Situational instructions and task order in recall for completed and interrupted tasks. *J. exp. Psychol.*, 1952, **44**, 434–437.

Henle, M. An experimental investigation of the dynamic and structural determinants of substitution. *Contr. psychol. Theory*, 1942, **2**(3).

Inglis, J. Abnormalities of motivation and "ego-functions." In H. J. Eysenck (Ed.), *Handbook of abnormal psychology.* New York: Basic Books, 1961.

Jourard, S. M. Ego strength and the recall of tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 51–58.

Kendler, Tracy. The effect of success and failure on the recall of tasks. *J. gen. Psychol.*, 1949, **41**, 79–87.

Lewis, A. B., & Franklin, M. An experimental study of the role of the ego in work: II. The significance of task orientation in work. *J. exp. Psychol.*, 1944, **34**, 195–215.

Lewis, H. B. An experimental study of the role of the ego in cooperative work. *J. exp. Psychol.*, 1944, **34**, 113–126.

McClelland, D. C. *Personality.* New York: Sloane, 1951.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

McConnell, T. R. Suggestibility in children as a function of chronological age. *J. abnorm. soc. Psychol.*, 1963, **67**, 286–290.

McKinney, F. Studies in the retention of interrupted learning activities. *J. comp. Psychol.*, 1935, **19**, 265–296.

Marrow, A. J. Goal tensions and recall. Part 2. *J. gen. Psychol.*, 1938, **19**, 37–64.

Martin, J. R. Reminiscence and gestalt theory. *Psychol. Monogr.*, 1940, **52**, 1–37.

Miller, M. B. "Rebelliousness" and repetition choice of a puzzle task under relative autonomy and control conditions in adolescent retardates. *Amer. J. ment. Defic.*, 1961, **66**, 428–434.

Nowlis, H. H. The influence of success and failure on the resumption of an interrupted task. *J. exp. Psychol.*, 1941, **28**, 304–325.

Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1951. Pp. 578–587.

Pachuri, A. R. A study of gestalt problems completed and interrupted tasks. *Brit. J. Psychol.*, 1935, **25**, 365–381, 447–457.

Rickers-Ovsiankina, Maria. Studies on the personality structure of schizophrenic individuals: II. Reaction to interrupted tasks. *J. gen. Psychol.*, 1937, **16**, 179–196.

Rosenzweig, E. The experimental study of repression. In H. Murray (Ed.), *Explorations in personality*. New York: Oxford Univer. Press, 1938. Pp. 422–491.

Rosenzweig, S. Preferences in the repetition of successful and unsuccessful activities as a function of age and personality. *J. genet. Psychol.,* 1933, **42**, 423–440.

Rosenzweig, S. Need-persistive and ego-defensive reactions to frustration as demonstrated by an experiment on repression. *Psychol. Rev.,* 1941, **48**, 347–349.

Rosenzweig, S. An experimental study of "repression" with special reference to need-persistence and ego-defensive reactions to frustration. *J. exp. Psychol.,* 1943, **32**, 64–74.

Rosenzweig, S. Further comparative data on repetition choice after success and failure as related to frustration tolerance. *J. genet. Psychol.,* 1945, **66**, 75–81.

Rosenzweig, S., & Mason, G. A. An experimental study in memory in relation to the theory of repression. *Brit. J. Psychol.,* 1934, **24**, 247–265.

Rosenzweig, S., & Sarason, S. An experimental study of the triadic hypothesis: Reaction to frustration, ego-defense, and hypnotizability. II. Correlational approach. *Charact. Pers.,* 1942, **11**, 1–19.

Sanford, N., & Risser, J. What are the conditions of self-defensive forgetting? *J. Pers.,* 1948, **17**, 244–260.

Sanford, R. N. Age as a factor in the recall of interrupted tasks. *Psychol. Rev.,* 1946, **53**, 234–240.

Sarason, S., & Rosenzweig, S. An experimental study of the triadic hypothesis: II. Thematic apperception approach. *Chart. Pers.,* 1942, **11**, 150–165.

Smock, C. D. Recall of interrupted and noninterrupted tasks as a function of experimentally induced anxiety and motivational relevance of the task stimuli. *J. Pers.,* 1956, **25**, 589–599.

Spradlin, J. Task resumption phenomena in mentally retarded Negro children. *Abstr. Peabody Stud. ment. Defic.,* 1955–60, **1**, 40.

Stedman, D. J. *Locus of control as a determinant of perceptual defense in mentally retarded subjects.* Ann Arbor, Mich.: University Microfilms, 1962.

Strother, G. B., & Cook, D. M. Neurocirculatory reactions and a group stress situation. *J. consult., Psychol.,* 1953, **17**, 267–268.

Tamkin, A. S. Selective recall in schizophrenia and its relation to ego strength. *J. abnorm. soc. Psychol.,* 1957, **55**, 345–349.

Winder, C. L. On the personality structure of schizophrenics. *J. abnorm. soc. Psychol.,* 1952, **47**, 86-100.

Zeigarnik, B. Das Behalten erledigter and unerledigter Handlungen. In K. Lewin (Ed.), Untersuchungen zer Handlungs: Und Affectpsychologie. *Psychol. Forsch.,* 1927, **9**, 1–85.

Zeller, A. F. An experimental analogue of repression: III. The effect of induced failure and success on memory measured by recall. *J. exp. Psychol.,* 1952, **42**, 32–38.

(Received January 2, 1964)

Psychological Bulletin
1964, Vol. 62, No. 5, 323–332

PRAISE AND BLAME AS INCENTIVES [1]

WALLACE A. KENNEDY AND HERMAN C. WILLCUTT

Florida State University

33 articles spanning 50 yrs. of research on the effects of praise and blame on the performance of school children are reviewed. Praise generally acts as a facilitator to performance, though often it is indistinguishable from practice effect. The single exception in recent years is the decrement in performance from praised underachievers. Exceptions to the generally decremental effect of blame on the performance of children are underachieving children, very bright adolescents, and Negro children working under Negro examiners. The problem of subject-reinforcement history is presented as a major contributor to the confusion in the literature, and the use of functional designs is suggested as a solution.

Experimental studies using various incentives for motivating discrimination, learning, and performance in human and animal subjects compose the greater part of the literature of psychology. The history of child psychology is spanned by studies seeking to evaluate the effectiveness of praise and blame as incentives for school children. The present review is confined to empirical studies considering the effects of the verbal incentives, praise and blame, on discrimination, learning, and motor tasks in school-age children below the college level. Philosophical and theoretical papers will be considered only when they are derived directly from empirical data.

Up to 1930

Eight years prior to Binet's monumental work on intelligence, Binet and Vaschide (1897), measuring the effects of verbal encouragement on the physical output of children, found that all 43 subjects improved their scores. In a later effort Kirby (1913) gave verbal encouragement, knowledge of results, a practice period, and then a retest to 1,350 third and fourth graders working arithmetic problems. Although all of the subjects achieved a gain, it is impossible to say whether the improvement was due to praise, knowledge of results, or the period of practice.

Although college students were utilized by Gilchrist (1916) and Gates and Rissland (1923), their work, which is therefore not pertinent to this review, must be mentioned because it provided the methodology for the later Hurlock studies which formed the beginning of the great surge of interest in the literature on the effectiveness of praise and blame as incentives for school children. Hurlock (1924) first studied the effects of verbal incentives on school children by dividing 408 third- fifth-, and eighth-grade children into groups by age, sex, race, and intelligence, administering an intelligence test in a test-retest design, a 1-week span between testings, and using praise, reproof, and control groups. Her overall conclusion was that neither praise or reproof was superior to the other as an incentive and that both tended to result in a greater improvement than did practice alone.

Further analysis of the different groups suggested that older children responded more to both praise and reproof

[1] The research reported herein was supported through the cooperative research program of the Office of Education; United States Department of Health, Education, and Welfare; Cooperative Research Project No. 1929.

than did the younger ones; that girls improved more from practice than did boys and were less influenced by either praise or reproof than were the boys; that Negro children reacted more favorably to praise, white children more to reproof. (This latter difference, however, was only slight, and, according to Hurlock, showed that the two races were more alike than different in respect to their reactions to these incentives.) Hurlock (1924) felt that some incentive was more essential for superior children, if their work was to be kept up to the maximum of their ability, than for inferior ones. Both praise and reproof seemed to serve this purpose. Reproof had its greatest effect on the children rated by their teachers as superior in school work; praise was of greater value for children rated as average and inferior.

Concerning the then current controversy over the constancy of the IQ, the next year Hurlock (1925a) followed up her work on intelligence test performance and again found no difference between the effects of praise and blame, both incentives raising the IQ significantly more than did practice alone.

Hurlock's (1925b) most distinguished work on praise and blame was an attempt to study these variables over a longer period of time.

> In a classroom, do the children who constantly receive praise for their work show more improvement from day to day than do the children who are reproved or who are completely ignored [p. 146]?

She used five modifications of the Courtis Research Test in Arithmetic over a period of 5 days, 106 children from Grades 4–6, the incentives administered in groups before the other members of the class, except for the isolated control group. Increased accuracy was found only in subjects who were praised or reproved, the greatest improvement being made by the praised group.

> These results as a whole point conclusively to the fact that regardless of the factors of age, sex, initial ability, or accuracy, praise is decidedly the most effective. . . . Reproof, when first used, seemed to be about equal in value to praise, but with continued use its effectiveness showed a decided decline [p. 159].

Two years later less difference was found between the groups, but the same trends were apparent when Cohen (1927) replicated the Hurlock (1925b) study.

A clinical interpretation was given by Bird (1927) to an analysis of factors influencing learning on a reading readiness test based on 100 children from 4 to 6 years old. Eight of the children were excessively dependent upon praise, four were more satisfied by admonition than indifference.

In the first major review of the literature in this area, Hurlock (1931) made no attempt to criticize or comment upon the past performance of researchers, or to point out new directions for later research to explore. In looking over the studies prior to 1930, confusion of results from research studying the effects of verbal incentives on children already is apparent. In the two earlier studies, verbal encouragement was found to improve the scores of all participating children on both the motor task of Binet and Vaschide (1897) and the regular classroom task of Kirby (1913). Neither study employed a control group or attempted to compare effects of praise versus reproof.

A very careful researcher, Hurlock, in two studies using an intelligence test as the task, one more or less a duplication of the other, found neither praise or reproof to be superior, although both were more effective than practice alone. In a third study using a regular classroom task she found praise to be decidedly more effective than reproof,

although both were still superior to practice alone. This latter finding was substantiated by Cohen's duplicate study. On a reading readiness test Bird found withheld praise caused delayed learning in some children, and reproof was necessary to encourage the best efforts of others.

None of these studies with differing results used the same age levels: Kirby used third and fourth graders; Hurlock's two earlier studies used Grades 3, 5, and 8, while her later study, and that of Cohen, used Grades 4 and 6; Bird used nursery and kindergarten children. Yet in the studies using multiple grades, no age differences were found. Could differences in methodology be responsible for such clashing differences in results?

1930 TO 1940

Beginning the work of the next decade, Warden and Cohen (1931) studied the effectiveness of schoolroom incentives on 38 fourth graders working arithmetic problems. They concluded that the commonly-used incentives of verbal praise and reproof were not as effective as might be supposed, that any change in the regular classroom routine worked as an incentive.

From their review of the literature, Davis and Ballard (1932) concluded that praise is better than reproof; that boys are more influenced by reproof, girls by praise; that, generally speaking, positive incentives are better than negative incentives for all ages, grades, and levels of intelligence.

Studying the relative effectiveness upon comparable groups of children of a number of different types of external incentives, Chase (1932) concluded that some motivation, whether positive or negative, is more effective than none. In a retest of 102 of Chase's original 213 children, Anderson and Smith (1933) concluded, in agreement with Chase, that some motivation is more effective than control motivation. There were differences between the two studies, however, in reference to the effectiveness of failure on the performance of the subjects, differences which might possibly be explained by the difference in age distribution between the two samples, since all of the original subjects could not be utilized in the follow-up study, and those used were older than they had been previously. Necessarily there were no younger subjects in the follow-up study, and older subjects were added.

The review by Brenner (1934) led him to conclude that the obvious conflict of results was bound to occur due to the varying experimental conditions and general inadequacies of the studies, but also "in a large measure to the faulty formulation of the problem [p. 48]." He felt that all the studies in the literature to date erred in assuming an absolute level of difference between incentives, which "act in specific situations, depending upon all the factors of the situation as a whole . . . and derive their attributes, so to speak, from the situation in which they are active [pp. 48–49]." Brenner (1934) stated that no general conclusions could be drawn regarding the effectiveness of praise and blame, but that their interaction with many other conditions needed to be considered.

> Proper application of the incentive in a given situation depends upon the insight of the teacher. The effectiveness or worth of a teacher depends upon his ability to make adequate use of motivation [p. 50].

Brenner himself investigated the effects of immediate versus delayed praise and blame and found no significant differences between the two.

Nor did Benton (1936), studying the influence of incentives on intelligence test scores, find any difference

between praise, reproof, or control motivation.

The first study, which attempted to take into account personality differences, was conducted by Forlano and Axelrod (1937) at the fifth-grade level. They concluded blame to be more effective than praise, but speculated as to whether more trials would have brought about significance for the praise groups.

The chief concern of Murphy, Murphy, and Newcomb (1937), who list and digest in a series of tables a large number of studies concerning incentives, is the problem of:

> The relationship between the individual personality and the situation. . . . All of the experimental work attempting to analyze the effects of praise and reproof . . . was until recently carried out without any references to the values or means of the situation' to the individual child and its relation to his own personality [pp. 449–450].

They call attention to the inconsistencies in the literature and attribute this confusion to a failure to consider the individual in relation to the cultural background and his own peculiar personality makeup.

Schmidt's (1941) concern is expressed at the conclusion of his excellent review of the literature:

> This review of the literature leaves no definite conclusions to be drawn as to the effectiveness of praise and blame as incentives to learning. This has been brought about, it is felt, through there being no clear-cut conception of the basic principles involved, or through a non-realization or non-differentiation of what was being sought and what was actually being done. It is true that performance has been mentioned frequently in the literature cited, but this has been done casually and interchangeably with learning [p. 7].

On one point all of the studies during the decade 1930 to 1940 agree: the experimental incentives were administered regardless of the actual performance of the children. This in itself is a somewhat artificial approach, since classroom teachers would supposedly be using the incentives to fit the occasion. But whether or not the answers were correct or incorrect, or the performance satisfactory or unsatisfactory, during this decade, as for previous pertinent research, subjects were administered the incentives at random.

Warden and Cohen (1931) not only found praise and reproof groups to be not significantly different from each other or the control group, but also to be inferior to promise of a game in motivating fourth graders on a regular classroom task. Brenner (1934) also found no differences between praise, blame, and control groups, even when divided into delayed and immediate praise and blame. Nor did Benton (1936) find any differences between praise and control: he did not make use of a reproved group.

Yet Chase (1932), using a motor task, found both praise and reproof to be better than control, and failure to be better even than praise. However, this latter finding was contaminated by previous incentive condition and method of administering the task. Anderson and Smith (1933), in a follow-up study, agreed with Chase's finding and definitely concluded blame to be more effective than praise. And Forlano and Axelrod (1937), endeavoring to take into consideration the personality of the subjects, found a significant difference only in the introverted subjects, where blame was more effective than praise or control. There were no differences in effectiveness of incentives for the extraverts.

None of these later studies agreed with Hurlock (1925b) that praise is more effective than reproof. The tendency for this period of research leans either toward no differences between praise and reproof, or toward reproof as being more effective. There is still no generally accepted conclusion from most

researchers in the field of effectiveness of verbal incentives for school children. It is no wonder that researchers are beginning to end their reports with statements to the effect that praise and blame must be used wisely and with understanding for the specific pupil at the specific time.

1940 to 1950

In an effort to study the effects of reproof, Potter (1943) concluded that at higher ages reproof had little effect on performance, while at lower age levels the effect varied.

Klugman (1944), making use of the equivalent forms provided by the 1937 revision of the Stanford-Binet, concluded that there was no significant reliability difference between praise and money as reward conditions; that there was a substantial superiority of the scores of Negro children tested with the money incentive over those tested with the praise incentive; and practically no difference between Negro and white children when money was the incentive; that white children showed a substantial superiority over Negro children when praise was the incentive. He (Klugman, 1944) accounted for the lack of significant difference between the incentives on the basis that the task was "so challenging that the use of one incentive rather than the other made little or no significant difference [p. 268]." When evaluating his results it must be taken into account that he did not make use of a control group.

Like Forlano and Axelrod (1937) earlier, Thompson and Hunnicutt (1944) were interested in the effects of praise and blame on the work achievement of introverts and extraverts. The methods of the two studies were similar with the latter studying its subjects over a longer period of time: six repeated measures under the reward conditions as opposed to three trials in the earlier study. Both the total praise and blame groups were significantly superior to the total control group, causing the authors to conclude that the effectiveness of verbal incentives in motivating pupils was more noticeable when introverts and extraverts were not differentiated. But there was no significant difference between the praise and blame groups.

Unlike Forlano and Axelrod, who found introvert blame to be significantly superior to introvert control and introvert praise, Thompson and Hunnicutt, for the first three trials, found no significant difference between these groups. On the final three trials, repeated praise increased the performance of the introvert-praise group until it was superior to introvert blame and extravert praise, answering in the affirmative Forlano and Axelrod's question as to whether repeated trials would raise the effects ot praise to significance. Repeated blame increased the performance of the extravert blame group until it was superior to extravert praise and introvert blame. On none of the tests did the extravert praise or introvert blame groups differ significantly.

Thus, for this study of fifth graders divided into introverted and extraverted groups, application of praise and blame, when repeated often enough, had differential effects: introverts achieved a higher level of performance when praised, extraverts when blamed. Reminiscent of Brenner (1934), the authors (Thompson & Hunnicutt, 1944) commented upon this conclusion by stating:

> The results of this study indicate that praise, as well as blame, can be used unwisely by the elementary school teacher if he does not fully appreciate and understand the different personalities present in his classroom. Praise and blame should not be judged on an either-or basis, but should be used to fit the case [p. 266].

Whether it was the effect of World War II, a shift in interest by investigators in the field, or high standards for

the next step set by Schmidt (1941), the decade 1940–1950 added only three studies to the investigation of the effects of praise and blame on the performance of school children. Only one of these, Thompson and Hunnicutt, endeavored to compare the effects of both praise and blame on student subjects.

Potter (1943), studying the effect of reproof on the performance of children at different age levels, found that it decreased performance of third graders, increased that of sixth graders, and had little effect on ninth and twelfth graders. In the first study not to employ a control group since that of Bird (1927), Klugman (1944), trying to establish a relationship between praise and money as incentives, found no differences in performance on the Stanford-Binet.

1950 TO 1960

In the 1950–1960 decade there began to appear in the literature a parallel line of investigations using knowledge of results as reinforcement, rather than praise and blame. There is, of course, a certain amount of overlap, but for reasons of clarity, simplicity, and emphasis, this series of studies will not be discussed.

To test whether praise was more effective when used in massed or spaced distributions, Mech, Kapos, Hurst, and Auble (1954) studied 240 fourth graders divided into massed versus spaced-reinforcement groups, further divided by 0%, 50%, and 100% verbal reinforcement. The massed-trained group performing significantly better, it was concluded that the intertrial interval is a factor determining the effectiveness of praise. It is unfortunate that in this study reward condition is confounded with percentage of reinforcement. Also there is the question of a fatigue factor. One group finished in approximately an hour, the other in 3 hours.

In his study to determine if the motivational needs of blind children are similar to those of sighted children, Kent (1957) hypothesized, from results of the Thompson and Hunnicutt (1944) study, that introverts would respond more readily to praise, extraverts to reproof, but found that either incentive was better than none for all groups except the public-school blind, where there was no difference between incentives and control. Praise was found to be more effective regardless of personality type or visual status.

Studying the effect of verbal incentives on reading improvement, Silverman (1957) concluded that teacher use of praise and reproof cannot predict the reading growth of pupils, but can predict total verbal output. A methodological disadvantage of this study, along with those of Potter and Brenner, was the use of intact classrooms and thus the inability to assign subjects randomly.

Terrell and Kennedy (1957), using a variety of reward conditions, rewarded and reproved their subjects according to their performance, and concluded that the candy-reward group had learned the task and had transposed this learning significantly more quickly than the other groups. Neither praise or reproof was significantly more effective than the control group.

Verbal encouragement and praise were concluded to have a definite facilitating effect upon the psychomotor performance of mental defectives by Ellis and Distephano (1959).

Studying the effect of praise and blame on a localization problem, Sandstrom and Weinz (1958) concluded that due to the conflict between encouragement and an unattainable reward the praised group's performance was inferior to that of the reproved group.

The decade 1950–1960 doubled the

number of studies of the previous decade and introduced interesting variations to the research theme. Blind children and mental defectives were studied; experimental incentives were administered according to the performance of the subjects; massed versus spaced praise was studied; and teacher use of praise and reproof in the classroom was recorded in an effort to evaluate its effect on reading improvement.

Results were just as varied, but almost predictable: Massed praise was found to be more effective than spaced praise; even for blind children praise and blame were more effective than control motivation, except for public-school blind where there was no difference between groups and where performance itself was significantly inferior to that of residential-school blind and public-school sighted children; teacher use of praise and reproof in a classroom cannot predict reading growth in pupils; candy is a significantly more effective reward than praise or reproof, neither of which is better than a knowledge-of-results control group; praise is significantly more effective for mental defectives than no incentive; reproof is better than praise if the reward offered is unattainable, although both are better than no information on results of performance.

Although Sandstrom and Weinz (1958) found reproof to be more effective, the conclusion was qualified by conditions in the praised group. No really radical findings are noted for this decade. Possibly its importance lies in the varied forms that the research was taking, in an evergrowing effort to try to reach some valid decision as to the effects of praise and reproof on the performance of school children.

1960 to Date

No significant differences were found between praise and control groups in the Dollins, Angelino, and Mech (1960) study on the effects of teacher praise upon 75 elementary school children scored below the thirtieth percentile on the California Test of Personality.

Stevenson and Snyder (1960), studying the use of praise and reproof with mentally-retarded children, established clear evidence that the effect of a particular incentive condition is influenced significantly by the preceding incentive condition: praise as the original incentive maintained higher performance even when followed by blame or no incentive, while with more praise performance was increased; blame followed by praise led to increased performance, but a decrement in performance was apparent for blame followed by no incentive. The authors concluded that when the examiner's statements led subjects to assume that a particular level of performance is expected or that his performance is less satisfactory than that of other subjects, failure increases motivation; but when the examiner's statements only comment upon subject's performance, failure lowers motivation.

Studying the relationship between concrete and abstract reinforcement on the sharing of preschool children, Fischer (1961–62), in a finding consistent with that of Terrell and Kennedy, concluded bubblegum as a reward was more effective than praise.

Metz (1961–62), studying the effects of stress and praise on creativity, concluded that stress facilitated performance of low scholastic-aptitude subjects, inhibited that of high scholastic-aptitude subjects, while praise facilitated the performance of all subjects.

Analyzing the effects of praise and blame as a function of intelligence, Kennedy, Turner, and Lindner (1962) studied two groups of adolescents with IQs from 124 to 150 and from 95 to 116. For the superior group no dif-

ferences in performance were found between the three incentive groups, while for the average group blame caused a significant decrement in performance.

To test the Kennedy, Turner, and Lindner findings on adolescents, Willcutt and Kennedy (1963) used the same procedure on fourth graders. At this grade level there was no significant relationship between intelligence and effectiveness of verbal incentive. There was a significant difference at this age, however, between incentives, praise being superior to either reproof or no incentive. One point to consider, however, in comparing these two studies, is that there were no students in the fourth-grade group with IQs comparable to the superior adolescents in the earlier study, the superior fourth graders being in the IQ range 111–130.

Changing the task from a discrimination problem to performance on an intelligence test, the grade level to second and third, and controlling for class and caste social groups rather than levels of intelligence, Tiber and Kennedy (1964), studying the effects of incentives on the intelligence test performance of different social groups, contrary to expectations, found no difference in effectiveness of incentives for any of the social groups, yet, as in previous research, found a significant difference in IQ scores between these social groups. The authors suggest that explanations of IQ differences between cultural groups must be based on causes other than lack of intrinsic motivation provided by the intelligence test itself.

Studying the effects of praise and blame on the paired-associate learning of 90 boys retarded and nonretarded academically (IQ means 107.2 and 108.8, respectively), Van De Riet (1963) found that blame was significantly more effective for the educationally retarded group, praise more effective for the noneducationally retarded group. She defended her hypothesis, that praise produced anxiety in the academically retarded group, with clinical observations.

Bluhm (1964), using the same discrimination task as Kennedy, Turner, and Lindner and Willcutt and Kennedy, studied the effect of anxiety upon the reaction to praise and blame of 472 fourth, seventh, and tenth graders divided by means of the discomfort relief quotient method described by Mowrer (1953), and found that reward condition does not interact significantly with anxiety level. Thus the main effect of the reward condition does not seem to be related to the level of anxiety.

Kennedy and Willcutt (1963) undertook a factoral study broad in scope and design. Their investigation of the effects of praise and blame on a discrimination task under the variables of grade, intelligence, sex, race, social class, school, and examiner concluded that, although the study came about due to evidence in the literature that the effects of praise and blame on subject performance was highly variable, for this study, using a discrimination task with speed as the criterion, the effects of praise, blame, and control were quite consistent regardless of subject differences: reproof had a debilitating effect upon performance, while praise resulted "in a somewhat higher increase than that coming from practice alone [p. 63]."

Using the same discrimination task, Vega (1964), following up the Kennedy and Willcutt (1963) study to test the effect of the race of the examiner on the reaction to praise and blame of 324 Negro second, sixth, and tenth graders, found that for the three white examiners the results were the same as in the original study. For the three Negro examiners, however, there was a significant difference, with all three incentive

conditions, praise, blame, and control, resulting in improvement in scores. The significant difference between the white and Negro examiners is based squarely on the subjects' reaction to blame. For the white examiners in both studies, blame caused a decrement in performance, even for the white subjects of the original study. However, for Negro subjects with Negro examiners, blame, as well as praise and no incentive, resulted in an increment in performance.

What have the 11 studies in the early part of this decade demonstrated about the effectiveness of praise and blame with school children? Anxiety seems to be unrelated to the effectiveness of verbal incentives, with the possible exception of the minority group of academically retarded children. Socioeconomic, school, and examiner variables seem not to be significantly related to the effectiveness of verbal incentives, with the exception of Negro subjects' reaction to blame under Negro examiners. Blame has been found generally to have a debilitating effect on the performance of school children. Praise has been found generally to have a facilitating effect on the performance of school children.

It now appears, then, that when one corrects for practice, as with the use of a control group, praise is a reasonably stable incentive, from study to study contributing an incremental effect upon the performance and learning of school children. An exception is the reaction of severe underachievers whom praise inhibits. Van De Riet (1963), in her explanation of these conflicting results, has suggested what is apparently the largest weakness of studies in this area, that thus far there has been very little control of the reinforcement history of the subjects, and that here a factoral design does not seem to be effective.

Blame, on the other hand, has fairly consistently exhibited an inhibiting effect upon the performance of school children. Exceptions to this general debilitating effect of blame are underachievers, Negro subjects performing under Negro examiners, and very bright adolescents.

Further clarification in this area would seem to await a functional design which in some way controls the reinforcement history of the subject. Using human subjects with such a complex phenomena as the effect of verbal praise and blame, control of the reinforcement history over a long span of time seems impossible, and there is some question as to the sufficiency of short-term control. Cross-cultural control of reinforcement history seems to raise more problems than it solves and would not appear to be any answer at all.

REFERENCES

Anderson, H. H., & Smith, Ruth S. Motivation of young children: The constancy of certain behavior patterns. *J. exp. Educ.*, 1933, 2, 138–160.

Benton, A. L. Influence of incentives upon intelligence test scores of school children. *J. genet. Psychol.*, 1936, 49, 494–497.

Binet, A., & Vaschide, E. Experiences de forces musculaire et de fond chez les jeunes garcons. *Année psychol.*, 1897, 4, 15–63.

Bird, Grace E. Personality factors in learning. *Personnel J.*, 1927, 6, 56–59.

Bluhm, P. M. Discrimination reaction time as a function of incentive-related DRQ anxiety and task difficulty. Unpublished doctoral dissertation, Florida State University, 1964.

Brenner, B. Effect of immediate and delayed praise and blame upon learning and recall. *Teachers Coll. Contr. Educ.*, 1934, No. 620.

Chase, Lucille E. Motivation of young children: An experimental study of the influence of certain types of external incentives upon the performance of a task. *U. Ia. Stud. child Welf.*, 1932, 5(3).

Cohen, A. A study of incentives under schoolroom conditions. Unpublished master's thesis, Columbia University, 1927.

Davis, R. A., & Ballard, C. R. Effectiveness of various types of classroom incentives. *Educ. Meth.*, 1932, 12, 134–145.

Dollins, J. G., Angelino, H., & Mech, E. V. With words of praise. *Elem. Sch. J.*, 1960, **60**, 446–450.

Ellis, N. R., & Distephano, M. K. Effects of verbal urging and praise upon rotary pursuit performance in mental defectives. *Amer. J. ment. Defic.*, 1959, **64**, 486–490.

Fischer, W. F. Sharing in preschool children as a function of amount and type of reinforcement. *Dissert. Abstr.*, 1961–62, 22:2871. (No. 61-5413)

Forlano, H., & Axelrod, C. The effect of repeated praise or blame on the performance of introverts and extroverts. *J. educ. Psychol.*, 1937, 28, 92–100.

Gates, Georgina S., & Rissland, Louise Q. The effect of encouragement and of discouragement upon performance. *J. educ. Psychol.*, 1923, **14**, 21–26.

Gilchrist, E. P. The extent to which praise and reproof affect a pupil's work. *Sch. Soc.*, 1916, **4**, 872–874.

Hurlock, Elizabeth B. The value of praise and reproof as incentives for children. *Arch. Psychol.*, 1924, No. 71.

Hurlock, Elizabeth B. The effects of incentives upon the constancy of the I.Q. *J. genet. Psychol.*, 1925, **32**, 422–434. (a)

Hurlock, Elizabeth B. An evaluation of certain incentives used in school work. *J. educ. Psychol.*, 1925, **16**, 145–159. (b)

Hurlock, Elizabeth B. The psychology of incentives. *J. soc. Psychol.*, 1931, **2**, 261–290.

Kennedy, W. A. Turner, A. J., & Lindner, R. Effectiveness of praise and blame as a function of intelligence. *Percept. mot. Skills*, 1962, **15**, 143–149.

Kennedy, W. A., & Willcutt, H. *Motivation of school children.* (Cooperative Research Project No. 1929) Tallahasee, Fla.: United States Department of Health, Education, and Welfare; Department of Education; October 1963.

Kent, H. R. The effect of repeated praise and blame on the work achievement of blind children. *Dissert. Abstr.*, 1957, **27**:675. (No. 20, 056)

Kirby, T. J. Practice in the case of school children. *Teachers Coll. Contr. Educ.*, 1913, No. 58.

Klugman, S. F. The effect of money incentive versus praise upon the reliability and obtained scores of the revised Stanford-Binet test. *J. gen. Psychol.*, 1944, **30**, 255–269.

Mech, E., Kapos, E., Hurst, F., & Auble, D. A mathematical description of the relationship between frequency of responding and intertrial interval in a motivational situation. *J. Psychol.*, 1954, **37**, 251–256.

Metz, W. A. The relative effects of stress and praise on creativity. *Dissert. Abstr.*, 1961–62, 22:2885. (No. 61-5649)

Mowrer, O. *Psychotherapy: Theory and research.* New York: Ronald Press, 1953.

Murphy, G., Murphy, Lois B., & Newcomb, T. M. *Experimental social psychology.* New York: Harper, 1937. Pp. 376–508.

Potter, E. H. The effects of reproof in relation to age in school children. *J. genet. Psychol.*, 1943, **63**, 247–258.

Sandstrom, C. I., & Weinz, E. Effects of praise and reproof in a localization experiment. *Acta psychol. Amsterdam*, 1958, **14**, 137–143.

Schmidt, H. O. The effects of praise and blame as incentives to learning. *Psychol. Monogr.*, 1941, **53**(3, Whole No. 240).

Silverman, H. F. Effects of praise and reproof on reading growth in a non-laboratory classroom setting. *J. educ. Psychol.*, 1957, **48**, 199–206.

Stevenson, H. W., & Snyder, Leila C. Performance as a function of the interaction of incentive conditions. *J. Pers.*, 1960, **28**, 1–11.

Terrell, G., Jr., & Kennedy, W. A. Discrimination learning and transposition in children as a function of the nature of the reward. *J. exp. Psychol.*, 1957, **53**, 257–260.

Thompson, G. G., & Hunnicutt, C. W. Effects of repeated praise or blame on the work achievements of introverts and extroverts. *J. educ. Psychol.*, 1944, **35**, 257–266.

Tiber, N., & Kennedy, W. A. The effects of incentives on the intelligence test performance of different social groups. *J. consult. Psychol.*, 1964, **28**, 187.

Van De Riet, Hani. The effects of praise and reproof on paired-associate learning in educationally retarded children. *Dissert. Abstr.*, 1963, **24**:1250. (No. 63-6367)

Vega, M. The performance of Negro children on an oddity discrimination task as a function of the race of the examiner and the type of verbal incentive used by the examiner. Unpublished doctoral dissertation, Florida State University, 1964.

Warden, C. J., & Cohen, A. A study of certain incentives applied under schoolroom conditions. *J. genet. Psychol.*, 1931, **39**, 320–327.

Willcutt, H. C., & Kennedy, W. A. Relation of intelligence to effectiveness of praise and reproof as reinforcers for fourth-graders. *Percept. mot. Skills*, 1963, **17**, 695–697.

(Received January 7, 1964)

Psychological Bulletin
1964, Vol. 62, No. 5, 333–347

EFFECTS OF DOUBLE STIMULATION: TEMPORARY INHIBITION OF RESPONSE[1]

DONALD REYNOLDS[2]

Michigan State University

This paper deals primarily with findings in the area of reaction time, with a lesser focus on dichotic listening studies and stereoscopic perception. The generality of a temporary inhibition of response (TIR) phenomenon, given double stimulation in close temporal contiguity, is posited. 3 alternatives are offered: an S-centered (information and/or "filter" theory), an O-centered (the "expectancy" position), and an R-centered (a competing-response position) explanation. The latter is suggested as most adequately dealing with the data, although the alternatives are not seen as mutually exclusive. A method for the training of prepotent responses is suggested, along with an associated line of research into the TIR phenomenon.

The phenomenon discussed in this paper is one which follows from the simultaneous, or nearly simultaneous, presentation of two stimuli to the subject. The behavioral manifestation of such stimulation, provided that the first and second stimulus are not identical (or nearly so), is a delay in response to one of them. In some cases, a delay may be found even when the second stimulus closely resembles the first. This phenomenon, herein labeled as "temporary inhibition of response (TIR)," is not the mere updating of the fact that a complex task requires more time to perform than a simple one. Note that a delay is posited to occur in only one of the two responses. Although the whole task (making both responses) often takes more time than the two component responses combined additively, the observation that only one of the two responses is delayed tends to bolster a competing response position and more adequately explains how the complex task takes more time.

The TIR is a useful concept because it permits operational specification of what may be expected to occur when two stimuli follow each other in close temporal contiguity; further, TIR provides a convenient point of departure for the comparison of different theoretical positions with respect to response output following such stimulation.

Explanations of delay in response have been advanced along traditional lines; these tend to be stimulus centered, organism centered, or response centered. The stimulus-centered position maintains that variation in response output is primarily related to variation in stimulus input. The organism-centered position would prefer to deal with the "expectancy" of the subject; anticipation, set, and expectancy are thought to have profound behavioral consequences. The response-centered position directs our attention to the competing response tendencies elicited by each of

[1] This paper was written while the author was in receipt of a National Institute of Mental Health Predoctoral Research Fellowship, Number MPM 16,194-02, from the United States Public Health Service.

The author wishes to acknowledge the helpful comments and encouragement of M. R. Denny and D. M. Johnson of Michigan State University. Development of the ideas contained herein are, of course, the sole responsibility of the author.

[2] Now at Training Methods Division, HumRRO, 300 North Washington Street, Alexandria, Virginia.

the stimuli in the double stimulation paradigm.

The present paper discusses several sets of assumptions and relevant findings in each of the following areas: stereoscopic perception; dichotic listening; and reaction time, with the focus of interest on reaction time studies.

The dichotic listening procedure involves the presentation of different stimulus-inputs to each ear of the subject, who is later required to report the contents of one message, or of both, or of selected components of one or both messages. Dichotic listening is dealt with herein in somewhat greater detail than is stereoscopic perception since workers in dichotic listening adopt a set of assumptions which lead to conclusions and inferences which represent one major attempt to handle the TIR phenomenon.

In this paper the reaction time studies reported are mainly those involving a "b-reaction," in which the subject reacts to the first stimulus with its experimentally associated response, and to the second stimulus with a second, usually different, response. Simple "a-reactions" are also discussed, in which the subject makes one response to a given stimulus, as the period between two a-reactions shares properties with the period between the presentation of the first and second stimulus in the b-reaction. This interval of time which separates the two stimuli is herein referred to as the interstimulus interval (ISI).

There are three aims of this paper: (*a*) To show, by convergent findings from three seemingly unrelated areas, that a TIR phenomenon exists. As indicated above, the TIR is posited to occur in situations in which the subject is presented with two stimuli in close temporal contiguity and required to respond to both of them. (*b*) To call attention to the fact that a response-centered explanation of TIR (or what used to be called delays in response to a second stimulus, e.g., the "psychological refractory period," which is herein assumed to be a special case of TIR) has in the past infrequently been seen in the literature. Such an explanation is important because either a stimulus-centered or an organism-centered explanation does not adequately account for the previous (and often divergent) findings in these areas. (*c*) To suggest a line of research by which a response-centered explanation of findings in the double-stimulation paradigm may be tested.

Stereoscopic Perception

It is known that presentation of views of the same, or very nearly the same, object in the right and left frames of the stereoscope leads to the phenomenon of binocular fusion, often with a perception of depth. When images quite disparate in content are so presented, for example, a red square to the left eye and a green one to the right, the "classical" binocular rivalry situation obtains; this involves a more or less rapid alternation of a single percept: red or green. Very infrequently there is a binocular fusion into yellow. The main difference between binocular rivalry and fusion antecedents seems to lie in the degree of physical disparity of the separate stimulus inputs. Although Bartley (1963) indicated that binocular rivalry and/or luster may be due to "the temporal organization of the impulses from the two eyes reaching a given cortical region [p. 51]," and although there are undoubtedly organism variables accounting for the rate of alternation, the prediction of binocular rivalry could be made independently of knowledge of these variables, given a stimulus input of a certain degree of disparity.

Another explanation of binocular rivalry is based on the expectancies of

the subject. Past experience with visual stimuli leads the subject to expect to receive identical, or nearly identical, stimulus inputs to the two eyes. When this condition is met, the subject fuses the two images as he has in the past; when this condition is not met, he engages in a "switching" process, in which some balancing of stimuli based on order of importance or unusualness (Berlyne, 1960) is perhaps made. If one of the stimulus inputs is physically dominant, then fixation of attention on that stimulus will occur often (Reynolds, 1962; Reynolds & Toch, in press). Woodworth and Schlosberg (1954) point out that fatigue or satiation may play a role in explaining why the percept shifts in this case (p. 78).

One could also examine the competing response tendencies involved in the reception and processing of two disparate stimuli. Much as the "important" or "unusual" stimulus would tend to dominate the binocular percept, so too might the stimulus which evokes or elicits a prepotent response. (The prepotent response may be considered as that response standing highest in a "habit-family hierarchy" associated with a given stimulus or stimuli. No Hullian endorsement is implied.) While making this distinction appears at first to be trivial and primarily semantic, it should be noted that the concepts of importance or unusualness of the stimulus require very exacting control of the stimulus configuration, if they are to be dealt with successfully in the experimental situation. A later section of this paper will outline a proposed method for the establishment of prepotent responses through training; this method would permit the operational definition of both a prepotent and of a competing response, thereby allowing a test of a response-centered explanation of TIR. Furthermore, since competing responses are posited to account for the alternation of images (in binocular rivalry), by differentially manipulating the strength of the competing responses one may predict rate of alternation. This is not possible when only the important stimulus is specified.

It should be noted that the above discussion carefully avoids review of physiological factors associated with binocular rivalry. Osgood (1953, pp. 259–260), after a review of physiological research, calls attention to the fact that there is anatomical separation of the two retinal fixation points; though there are no direct (and few indirect) provisions for the fixation points to work as a unit, they do so. It is tempting to conclude that visual integration (fusion) is under some higher, central control. If so, then learning may be posited as mediating binocular fusion. Berkeley, in the eighteenth century, advanced a similar view applying to visual space perception. Need we develop a different process for binocular rivalry than for fusion? If rivalry is learned or subject to the laws of learning then a competing response explanation of it is not untenable.

The main point of interest in the binocular rivalry situation is, for the writer's purposes, this point: Even though stimuli are presented simultaneously to the subject, he responds to them successively. This is not due to some verbal impediment resulting from the inability of the subject to say red and green simultaneously, but reflects a genuine perceptual process which involves a temporary inhibition of response.

Dichotic Listening

In this area there is no single phenomenon as well defined as binocular rivalry in stereoscopic perception. There are certain patterns of delay of response which occur in highly reliable fashion—given certain stimulus conditions. Allan

(1961) has suggested that when "too much" stimulation for "immediate absorption" is presented to the subject, a perceptual organizing activity occurs which is absent when the subject is operating "below capacity." This notion, aside from being intuitively satisfying, seems to be able to explain "grouping," the making of what appears to be a single unified response consisting of two response units to two temporally close stimuli. A disadvantage of this notion is that it tends to be a circular one; the definition of below capacity hinges on whether or not "organization" occurs.

Broadbent's theoretical paper (1957b) is of interest here. In order to conceptualize the results of many dichotic listening studies he develops a model which takes the shape of a Y shaped funnel, down which balls (representing information) can be dropped. Balls dropped simultaneously down each branch will collide, and either remain at the fork or emerge successively. "Simultaneous stimuli either jam or produce successive responses [p. 208]." If this formulation is correct, there should be a TIR when stimuli are temporally close or arrive together. There is one exception to this rule, however: when the two stimuli are of low information content they may pass through and be processed together. There seems to be no contradiction between Broadbent (1957b) and Allan (1961); when the stimuli make fairly low demands on the subject, two or more may be dealt with together. Additional references are Broadbent (1954, 1956) and Bryden (1962).

Of import in dichotic listening studies are error rates and order of recall. It has been found that, when information is of diverse content, auditory fusion is extremely rare in the dichotic situation. If, for example, we present the stimuli 3-4-5 to the right ear and 6-7-8 to the left, the subject does not report the sequence 3-6-4-7-5-8 (unless a slow temporal order is used; see Bryden, 1962), or complain that he could not hear the stimulus input ("masking"), but typically reports all of the content from one ear, then all of the content from the other. Broadbent (1957a, 1958) thinks that stimuli from one pattern are held in "short-term storage" while stimuli from the other pattern are organized or otherwise dealt with (1958, pp. 298–299). Listening to two sets of digits in this situation leads to more errors from the "channel" which reports second (Broadbent, 1957b).

Bryden (1962) found that: (*a*) When the input of dichotic material is of slow tempo (one pair of stimuli every 2 seconds), reports tend to follow the temporal order of the stimuli, with some evidence of rehearsal between pairs of stimuli. (*b*) Although the longer lists led to more errors, length of list had no consistent effect on order of recall. (*c*) Speeding up the tempo (to one pair of stimuli every .5 second) leads to a breakdown of the "ear order" of recall. (The connection between this ISI and the range of ISIs in which the psychological refractory period is found in reaction time may be coincidental. On the other hand, an ISI of .5 second may represent a boundary between different ways of handling stimulus input.) Thus the ear order seems to be a function of the temporal spacing of stimuli, rather than of length of message (and, by implication, of storage capacity).

In the shadowing of dichotic messages, one of which was labeled by the experimenter as "irrelevant," Moray (1959) found that the subject was unable to report any content of the "rejected" message. Only the subject's own name seemed to "get through." Other types of irrelevant messages, including a pure tone (usually perceived as such), a man's voice later switched

to that of a woman, speech in a foreign language, and reversed speech were not clearly identified; if the subject was trying to "assist" the experimenter by failure to report content of the rejected message, it nevertheless seems likely that he would be able to report on the formal or structural properties of the stimulus input.

There are at least two ways to explain why the subject responds to his own name (assuming identification is a repsonse); each of these ways also explains why the rejected message fails to get through. If we make the assumption that the occurrence of the subject's name in this situation is a rare event of low probability, the occurrence of the name conveys considerable information, and thus may be attended to. When we try to apply this explanation to why the subject fails to get the irrelevant message (presumably also of fairly low probability in this situation), the analogy breaks down. Another way of looking at this is to assume that a response to one's own name is more easily elicitable than many other responses, that is, that the stimulus of one's own name is a potent elicitor of a response. Applying this assumption to the failure of irrelevancy to get through, we can conceive of the possibility that the irrelevant message (or stimulus) is not a potent elicitor of a response; in fact, the experimenter has told the subject to disregard that part of the stimulus input!

Treisman (1960) instructed the subject to "shadow" one of two messages presented dichotically. In the middle of the passage (connected verbal discourse), stimulus inputs were switched to the opposite ear without informing the subject. Typically, the subject would repeat one or two words from the now "wrong" ear, but never more than this. Aside from the contextual and meaningfulness variables apparent here, one must also note the latency of the correct response. A delay of response here seems a function of the replacement of one response by another ("wrong"), followed by an amendment of the incorrect response. A much more severe case of this can be illustrated by pointing out the difficulty subjects sometimes have in "breaking set"; set can be conceived as a series of prepotent responses established through prior experience.

A pattern seems to emerge from these dichotic listening studies: When the subject is presented with an "easy" situation (e.g., slow tempo of events) the response pattern flows smoothly with few delays or errors in response. At "harder" levels the subject either attempts some organization of stimulus inputs (e.g., ear ordering), or some selection of response output based on the potency of the eliciting stimulus pattern; in both cases delays in response (TIR) are involved, almost by definition.

Berlyne's statement (1960) seems to support a response-centered explanation for the findings in dichotic listening at these harder levels.

> If he [the subject] has an innate tendency to react to a complex of stimuli in a particular way, or if he has learned to perform a particular response in the presence of the whole complex, we had best talk about the selection of a response to the situation [p. 7].

He does not exclude stimulus selection, however, in such cases. It seems to be a case of part versus whole learning, with the latter favoring response selection. Since the subject is thought to respond in a particular fashion to the stimulus complex (by "organization of a response"), some response competition may be fairly hypothesized.

It will be recalled that Broadbent's filter theory, which deals with organizational capacities of the subject, is

mainly a stimulus-centered approach. He speaks of stimuli arriving simultaneously or successively, of having much or little information content, of being more or less similar to other stimuli, and of being highly probable or improbable (Broadbent, 1957b, 1958, pp. 298–300). All of these considerations are admittedly important. However, one discovers that the amount, latency, or correctness of response are secondary considerations taken to support the existence of the posited filter. Stated in another way, Broadbent has examined a variety of stimulus inputs and their consequent response outputs in an attempt to account for what might be going on inside the subject and has arrived at a position requiring selection of stimulus input as the major variable. Berlyne (1960, pp. 7–8) has pointed out that only some of the time should this be so.

One of the deficiencies of dichotic listening studies for the examination of the effects of temporally close double stimulation (of which simultaneous stimulation is the very obviously limiting case) is in the requirement of verbal responses by the subject. The subject cannot make simultaneous verbal responses, only successive ones. It will be seen that the reaction-time paradigm avoids this difficulty. One way of converting the state of affairs found in dichotic listening tasks into an asset would be to posit that the prepotent response will emerge first, given double stimulation. One could then attempt to test this by the establishment of response strengths by training responses to stimuli to which they would never be made "in nature." One would then provide double stimulation and examine response output in terms of latency and/or correctness of response. A further discussion of this training method will be deferred to a later section of this paper.

Reaction Time

As the interval between two stimuli becomes very short the reaction time to the second stimulus presented becomes lengthened as ISI is shortened and approaches .5 second (Adams, 1961, p. 68). This delay in response to the second stimulus is known as the "psychological refractory period," and may be considered as an analogue to binocular rivalry in stereoscopic perception. The remainder of this section is devoted to findings bearing on the psychological refractory period, and theoretical problems posed by it. The psychological refractory period is assumed to be a special case of TIR.

As far back as 1931 Telford found that the longest reaction times were to .5-second ISIs, when using a simple a-reaction task. Although he reported no statistics, he concluded:

> immediately after responding . . . there is a period of intrinsic unreadiness for response as shown by the lengthened reaction times to stimuli applied during this interval [p. 12].

Vince (1948) was one of the first to use a b-reaction, while varying the ISIs from 50 to 1,600 milliseconds. She too reported no statistics, but the general finding was that with ISIs of 50–500 milliseconds there was a significant delay; at greater ISIs there was no delay. She concluded that a "central computing process" may need 500 milliseconds due to "psychological refractoriness" and tried to tie this to other tracking tasks in which 500 milliseconds appears necessary to make compensatory movements. In this study the psychological refractory period is unfortunately confounded with the motor time needed to correct a response, as Poulton (1950) points out. If a second signal to respond occurs while the pencil is tracing the first response (as it did in her study), this latter motor response must be stopped and a new one started. Poulton believes that when a proper

preparatory set is permitted to form, the psychological refractory period may reduce to an interval of 200–400 milliseconds.

The role of preparatory set in reaction time using a b-reaction was also dealt with by Elithorn and Lawrence (1955). They varied the ISI (0, 20, 50, 100, 250, 370, 500, 600, 750, and 950 milliseconds) in a quasi-random procedure, finding that the greatest delay to the second stimulus was at the 100-millisecond ISI. There was much intersubject variability.

Using four practiced subjects with about 15,000 observations each, Hyman (1953) found that reaction time tends to be proportional to the amount of information in a signal. An interesting finding was that:

> In both Exp. II and III [in which there were two or more components with different amounts of information associated with each] the components within a condition interacted with each other . . . the reaction time to the low information component was higher, and the reaction time to the high information component was markedly lower, than would be predicted on the basis of the regression line fitted to the means of the conditions. This effect was very marked and occurred without a single exception for all four *S*s in both Exp. II and III [p. 194].

For example, predicted mean reaction times for .30 and 4.00 bits of information per signal were 258 and 824 milliseconds, respectively. The obtained mean reaction times were 306 and 585 milliseconds, respectively. When these separate mean reaction times are properly weighted and averaged, the obtained grand (condition) mean was 361 milliseconds, which closely compares with the predicted 363 milliseconds.

Note that the "common event"—the one with a higher probability of occurrence—has a longer reaction time than would be predicted from information theory calculations alone. It is not completely unlikely that common events are those to which there are many competing responses. A relatively "rare event" might have fewer competing responses, and thus reaction time to it be lowered accordingly. If increasing the amount of information per signal leads to a nonmonotonic reduction of competing responses, this might explain the greater discrepancy between expected and obtained reaction times to the rare event.

The applicability of Hyman's (1953) study to the psychological refractory period flows from the fact that "speed" and "load" are complementary functions in information theory. Shortening the ISI should be the same, or nearly so, as increasing the load. Hyman found that increasing the load leads to slower reaction times, and so should shortening the ISI. (Unfortunately, Hyman did not manipulate ISIs.) This is generally what happens, up to the point at which the ISI is short enough so that the double stimulation can be reacted to as if it were single. If speed and load are truly comparable in this situation, then the competing response explanation outlined above may be provisionally considered as tenable in dealing with the effects of short ISIs in conjunction with double stimulation.

Is the psychological refractory period confined to stimuli from the same mode, or is it general across modalities? To answer this question Davis (1956) used two visual stimuli and a visual-auditory pair (Davis, 1957), while varying ISIs from 50 to 400 milliseconds (1956) or 500 milliseconds (1957). The pattern of results is the same in unisensory and bisensory presentation; only at the shorter ISIs did delays of response occur to the second stimulus. In a follow-up study Davis (1959) hypothesized that it is the paying of attention to a signal, rather than making an overt response to it, that leads to greater delays in reaction time to the second stimulus. His results were quite com-

plex, but were taken to show support for a hypothesized "common analyzer or classifying system." Although he did not present statistics to support his conclusion, computation of t tests from his presented data reveals that there is no significant difference between mean reaction times in a unisensory condition in which the subject either did or did not respond to the first stimulus ($p > .05$, two-tailed), but a highly significant difference when a similar comparison is made in a bisensory condition ($p < .0001$, two-tailed).

The unisensory versus bisensory finding is tantalizing. If both stimuli come in on the same modality, then whether one makes a response to both or only to the second seems to make no difference in latency of the second, since the second reaction in both cases takes about the same time. But note what happens when the second modality stimulated is different from the first: the second reaction may take longer if the subject has to respond to both stimuli but not if he is to respond only to the second. In view of Broadbent's (1956) findings in dichotic listening with bisensory stimulation, one would expect the unisensory and bisensory stimulation to result in roughly comparable findings. That this is not the case might prove difficult for filter theory to handle. In summation: (*a*) Davis' conclusions are not quite supported by his own data, as there are discriminable differences between the unisensory and bisensory conditions; (*b*) evidence supportive of a "common or central analyzer system" has not been compellingly set forth; (*c*) and making of an overt response is not quite as unimportant as suggested by Davis, given the findings in the bisensory condition.

One of the first to inquire into the temporal spacing of the double stimulation was Klemmer (1956), who studied the effects of "time uncertainty" on simple reaction-time tasks. Using five different ISIs of 1–12 seconds in an a-reaction with a variable preparatory interval, he found that reaction time tends to increase with an increase in preparatory interval variability. Since the event (the stimulus to be responded to) was certain, the increase in reaction time was thought to be due to the temporal uncertainty associated with variable preparatory intervals. A series of recent studies have further explored the effects of time and event uncertainty.

Using almost exactly the same procedure as Davis (1959), but with fixed ISIs for each series, Borger (1963) presented two stimuli to the subject (auditory, then visual or vice versa) and varied the ISI from 50 to 500 milliseconds. The second response was found to be slower when the subject had also to make a first response, but not so when the second response was the only response made. This finding held only for ISIs of 300 milliseconds or less. The subjects were very variable and statistics were not presented. Borger felt that a theory based on time uncertainty was not supported by this data, since all ISIs were constant within a series. He brought out a good point: responses may be stored, as well as signals. With successive stimuli the subject may store the first signal until after the second has been received and then respond with a joint response to both. He may also decide on a response and store the response until the second signal has been received. As of now, it is not certain which (or if both) goes on (Borger, 1963, p. 8). This explanation would seem to dictate, however, that it is the first response, not the second, which is delayed in either of the two cases above.

Although Borger's (1963) study is generally in line with Davis (1959), it is not in agreement with Klemmer

(1956) with respect to time uncertainty as a factor in reaction time. Since Klemmer used a rather long ISI compared to Borger, and furthermore required an a-reaction rather than Borger's b-reaction, the studies are not strictly comparable. The expectancy of the subject may be a more potent variable when the preparatory intervals vary across a wide range of some considerable magniture (1–12 seconds), as compared to when the ISI has a possible range of 50–500 milliseconds.

Adams (1962) varied a time uncertainty condition associated with a 100-millisecond ISI, with an ISI range of 0–800 milliseconds. He found that, at less than a 200-millisecond ISI, the bisensory condition led to a delay in response to the second stimulus—the psychological refractory period. At 400 milliseconds or more, there is little evidence of the psychological refractory period, although some errors of anticipation occur. These results were taken to support an expectancy position; the subject was thought to be less than optimally prepared at short ISIs. Using a similar setup, with all ISIs = 0 milliseconds, Adams and Chambers (1962) had the stimulus events certain or not, with a unisensory or bisensory task. The bisensory group was superior when the stimulus events were certain, and this was thought due to increased speed associated with anticipation. When events were not certain the normally faster auditory reaction time tended to slow and eventually synchronize with the usually slower visual reaction time in the bisensory group.

The above two studies, when taken together, seem to show that the net effect of time uncertainty is to cause a slowing of the second response only at shorter ISIs ("psychological refractoriness"), whereas the effect of event uncertainty is to generally depress performance. Creamer (1963) addressed himself to this problem. He assigned five groups of subjects to time-certain conditions, in which the ISI used was homogeneous for the group. A sixth group (time uncertain) had five different ISIs in a random order; all groups were presented with two uncertain visual or auditory stimuli. The time-uncertain group was found to have a slower "response speed" (this includes errors of anticipation) than the time-certain group, but this difference was not statistically significant. Upon closer analysis, the time-uncertain variable was found to prolong the effects of the event-uncertain variable, resulting in the delay of response speed to the shorter ISIs. Since the time-certain and uncertain conditions were not different from each other on any comparison, Creamer concluded that it is event, not time uncertainty which is responsible for the psychological refractory period. Response delays were ascribed to a "human limitation for the central decision functions of processing choices among stimuli [Creamer, 1963, p. 194]." It may be tenable to assume that when events are uncertain (given a simple two-choice situation), responses to both must be held in readiness to insure maximum efficiency of response to both. When either stimulus occurs, there well may be competing response tendencies which are associated with the other stimulus which interfere with the response to the first stimulus. If so, the first response may be delayed. This is not usually the case, since it is the second response which is often delayed. Let us consider the implications of assuming equality of response strength: if one of the two responses available to the subject is not stronger than the other, then the response elicited by the first stimulus may become the prepotent response. If this is true, we may expect the second response to be delayed if and only if the second stimulus occurs

"too soon." At ISIs greater than 500 milliseconds, whatever effect the making of the first response has on the making of the second appears to have dissipated. This notion is further discussed in the next section of this paper.

Davis (1962) had two visual signals, each with a probability of occurrence = .5. He studied the relationship between the reaction time to the second stimulus and the ISI when: (*a*) The subject was told to ignore the first signal. (*b*) The subject had to respond on the nature of the first signal (while not responding to it) after response to the second signal. The ISIs were 50–250 milliseconds in steps of 50 milliseconds; a regular and a random series of ISIs were used. The findings were: (*a*) Only in the random series of ISIs was there an apparent delay in response to the second stimulus, and this occurred only at the 50-millisecond ISI. (*b*) Reporting on the nature of the first signal did not consistently increase reaction time to the second. This is another piece of evidence against Davis' belief that paying attention to the first signal, rather than the making of a response to it, is what leads to delays in reaction to the second.

Davis (1962) explains the delay at the short ISI as due to random presentation of the ISIs; the subject is thought to begin analyzing the first signal before the second signal is in, and then this leads to blocking. This seems ad hoc and does not explain why blocking, or delay, in the second response should not occur routinely. If the ISIs were longer, it is still more certain that the first signal will be analyzed before the second; why not greater delays in response to the second stimulus at longer ISIs? An alternate explanation is that short ISIs, and especially irregular ones with equiprobable stimuli involved, lead to maximum competing responses. Berlyne links competing response tendencies to delays in reaction time (1957, p. 334).

In order to compare reaction times to single and to first stimuli, Gottsdanker, Broadbent, and Van Sant (1963) used several tasks as follows: single-choice, left hand; single-choice, right hand; double-choice, left hand first; double-choice, right hand first. For each subject the mean reaction time for the first response was shorter in the single-choice situation. The authors (Gottsdanker et al., 1963) conclude:

> Lengthening of reaction time occurred because *S* had a subsequent task to perform with a different response unit. The two movements became part of a unified complex *solely because* they were channeled through a single operator [p. 167; italics added].

In the studies of Davis (1956, 1957, 1959) reaction time did not appear longer at 50 milliseconds than at 400 or 500 milliseconds for the first response. (Vince, 1948, had some evidence for the "grouping" of responses at or near a 50-millisecond ISI.) Gottsdanker et al. mention in a footnote (p. 165), that the second response seems to be shorter than the first by about 25 milliseconds. It would appear that some grouping of response is occurring in the Gottsdanker et al. study such that speed of the second response is either facilitated or at least not retarded. Although one might wish to use this study to claim support for an "expectancy" position, the fact that Davis did not find the first response delayed counsels caution in accepting an interpretation based either on anticipation (in which the first response need not be delayed, but the second may be facilitated, as in the grouping of responses) or on a conception of the subject as a single channel (in which the first response may be delayed but the second need not be) in the double-choice situation.

Will there be grouping at quite short ISIs? Halliday, Kerr, and Elithorn (1960) presented paired stimuli with

ISIs of 50 and 100 milliseconds, and single stimuli, 0, 50, and 100 milliseconds after a warning signal. In 25% of the presentations the second signal was omitted; this constituted the single presentation. Here we have both time and event uncertainty with very short ISIs. The general finding was that a large number of second responses were not delayed but were as fast as the first responses. ". . . the high expectancy of S_2 [the second stimulus] may account for both the shortness of delays . . . and their disappearance with training [Halliday et al., 1960, p. 88]." The first response did not seem delayed in this study. It seems somewhat curious that an expectancy position was considered tenable, and that the second stimulus was assumed to be highly expected, when in fact 25% of the time it was absent. If anything, one would have predicted that a large number of first responses would have been delayed (if expectancy were operative) while the subject waits for a possible, but not certain, second stimulus.

An interesting finding was recently reported by Helson and Steger (1962). A first response was found to be delayed when a second signal followed the first stimulus by 10–170 milliseconds, with a maximum effect in the range of 40–140 milliseconds. In this study the second signal was not considered a stimulus, since no response to it was required of the subject. This finding must give one pause: If the subject is not required to respond to both stimuli, how is it that his only response is delayed by a second, presumably irrelevant, signal? Here one sees the inadequacy of a stimulus-centered, or organism-centered, explanation of these findings. Instead, one might wish to consider what a competing-response position has to contribute.

Let us assume a mean reaction time close to 200 milliseconds for each subject, and let us further assume that a response "unfolds" or develops over time. One may then conceptualize a graph of response segments plotted against unit time (on the abcissa). If we assume the simplest (but not necessarily correct function) of a straight line, then each increment of 20 milliseconds should result in very roughly 10% of the response as it moves toward completion. If the maximum effect of the application of a second signal is in the range of 40–140 milliseconds, then we can see that the response is somewhere between 20% and 70% complete when the second (irrelevant) signal is applied. It seems likely that a positively accelerated curve more accurately depicts response segments as a function of unit time. If so, the above discussion must be modified. The curve will be displaced to the right of the straight-line function, indicating that relatively more time must pass before a response is, say, 50% complete. Thus the effect of the formation of competing responses may be magnified, and the more rapid the acceleration of the curve the greater the effect. Two empirical questions remain: Will competing responses form to a signal which is not to be responded to? Does enough time elapse during the elaboration of a response (such as in reaction-time experiments) for a competing response to form? One may tentatively conclude that not enough evidence has been marshaled to answer both questions in the negative.

A second study was recently reported by Helson (1964); in brief, his results were that when a tone is the second signal and a light is the "primary" signal (or vice versa), the following occurs: (*a*) At an interval of 0–25 milliseconds after the primary stimulus, the second signal (which is heteromodal) acts to facilitate or speed up the reaction. (*b*) At 25–35 milliseconds, the second signal appears to have no effect on the speed of the first (and

only) reaction—this is Helson's "neutral zone." (*c*) At greater than 50 milliseconds the second signal acts to slow up the first response. Can competing-response theory account for this paradox? If we recall the results of the study of Vince (1948), who found grouping of responses at ISIs of at or near 50 milliseconds, we may be able to approach an explanation of this paradox. If at 0–25 milliseconds there is a summation effect, that is, the double stimulation is perceived as single via grouping, then the net effect is perhaps analogous to the receipt of a single, more intense stimulus. It is known that reaction time is decreased as the intensity of the stimulus is increased (Woodworth & Schlosberg, 1954). This would tend to result in quickened reactions. If competing responses can adequately explain the lengthening of reaction time at intervals greater than 50 milliseconds, then the interval of 25–35 milliseconds (in which no effect is apparent) may show the combined effects of grouping and competing responses, that is, mutual cancellation. It should be underlined that this way of conceptualizing the data has, as yet, no empirical support.

In a study which may throw some light on the potential of a competing response theory, Morin and Forrin (1963) held transmitted information constant, finding that reaction time was positively related to "response uncertainty." This response uncertainty was introduced by permitting the subject to "randomly select" one of several correct responses to a single stimulus with which they had been experimentally associated. Morin and Forrin point out that the proportion of variance in reaction time accounted for by response uncertainty was twice that accounted for by transmitted information in their study (p. 34). Although this finding in no way invalidates the importance of transmitted information in reaction time studies, it should sensitize us to the possibility of competing-response effects in such studies.

Discussion

The area of reaction time studies in which the psychological refractory period is examined is one in which there is a welter of conflicting findings and general incompatibility across studies. The British studies typically use two to six subjects and many measurements. Conclusions evolve from inspection of the results, rather than the testing of hypotheses. In most of these studies it is common practice to omit reports of statistical tests, if in fact they were performed. Generalization from the British to the American studies and vice versa is tenuous, for the former use well-trained subjects and the latter typically do not (though making up for this, in part, by using larger numbers of subjects per group). The American studies tend to be "cleaner" but variations in "set" are hopelessly confounded across subjects.

Welford (1952) sought to explain the psychological refractory period by citing three classes of previous theorization: (*a*) The first stimulus renders the response mechanism insensitive so that it reacts more slowly to the second stimulus, especially at short ISIs (the "central delay" theory). (*b*) The delay in response to the second stimulus is due to its unexpected arrival (the "expectancy" position). (*c*) The delay is due to some phase of the two responses not being able to overlap (a "competing response" position?). After reviewing the available literature, he concluded: (*a*) A central organizing time is needed which is equal to the reaction time for a single response; regardless of the length of the ISI this time remains relatively constant. (*b*) Feedback from the first response may engage the cen-

tral mechanisms, and the perception of this feedback will require time. (*c*) No two central organizing times can overlap, so information from the second stimulus must wait in storage if information from the first stimulus is being dealt with at that time. His conclusions tend to be organism centered; unlike Broadbent, he does not unduly stress the stimulus content.

Other theories (stimulus centered or organism centered) are not incompatible with a competing response theory. The difficulty with a stimulus-centered explanation is in determining what guides stimulus-selection in, for example, the steroscopic perception or dichotic listening areas. The organism-centered explanation, in relying on the expectancy of the subject (or his property of being a single channel through which information passes), is in danger of committing a nominalistic fallacy; it labels, rather than explains. If expectancy of the subject could be somehow quantified and rigorously varied, one would feel more comfortable with the position, since specific operations could be produced as referents.

The chief advantage in dealing with the TIR phenomenon by means of a competing response position is that operational referents for the strength of a response tendency may be provided. If we can assume, for the moment, that double stimulation elicits competing responses in the subject, we may expect the TIR for a short time, after which the prepotent response should emerge. By training responses to specific stimuli previously unassociated with them, we may operationally define a prepotent response. Such definition is crucial in studying the psychological refractory period. All previous work tends to implicitly assume equality of response strengths, that is, the absence of a prepotent response.

As mentioned in the discussion of Creamer's (1963) study, if one of the two responses available to the subject is not stronger than the other, then the response elicited by the first stimulus may become the prepotent response, with predictable delay to the second stimulus if it arrives too soon. But what if the first response is relatively weaker than the second? Would the first response be delayed and the second accelerated? Would this result in a breakdown of the psychological refractory period? It is not inconceivable that some of the diverse results with respect to speeds of the first and second responses, relative to each other, may be due to unequal response strengths.

What we have been talking about all along is a "conflict" situation. Berlyne (1960) lists four factors influencing the degree of conflict:

> [*a*] the nearness to equality in strength of the competing response tendencies, [*b*] the absolute strength of the competing response tendencies . . . [*c*] the number of competing response tendencies . . . [*d*] the degree of incompatibility between competing response tendencies [pp. 32–33].

As for Factor *b* above, if the probability of the first response plus that of the second is greater than 1.0, the responses may not be considered wholly incompatible. Further, in the double-stimulation paradigm, the number of competing response tendencies (Factor *c* above) may be considered relatively controlled across subjects, if a *b*-reaction is required of the subject. Since separate response keys are usually provided for the subject in such paradigms, the fourth consideration above does not seem to be a potent variable. Thus the major factors to focus on are: the degree to which the first and second responses approach each other (nearness to equality in strength), and the absolute strength of the first versus the second response (provided that the sum of their probabilities does not exceed 1.0).

Suppose that we are able to succeed in training responses such that we can specify their probability, given a particular stimulus complex. We should therefore expect the greatest delay in the first response made given the situation in which the probability of the first response (R_1) equals that of the second (R_2), both of which equal .5. In this case, given an ISI of 0 milliseconds, we should expect R_1 to precede R_2 about half the time, and half the time to follow it. If the probability of R_1 is greater than that of R_2, we should expect R_1 to precede R_2 most of the time at an ISI of 0 milliseconds. Furthermore the probability of both responses may be differentially varied (e.g., with one equal to .8 and the other to .2) when the ISI is not equal to 0 milliseconds. This should have differential effects on the speeds of both responses along the lines barely suggested above.

Suggested Training Method

An interesting finding has come from the probability-learning literature. It is that subjects, in a probability-learning task, typically do not use a "maximum-success strategy," but tend to follow fairly closely the objective probabilities (Restle, 1961). Given a two-choice probability-learning situation, with a probability of the first stimulus of .7, and that of the second of .3, subjects do not always choose the first stimulus, but tend to distribute their responses about .73 and .27, respectively. Note that a maximum-success strategy, although it may operationally define a prepotent response (as outlined below), will not permit a competing response to form.

The probability-learning paradigm could be selected to train a prepotent response and one competing response in the double-stimulation task. This would permit study of a competing-response position by utilizing the TIR phenomenon. The subject would be placed in a situation in which he is to guess which of two lights will next be turned on by pressing the associated key; this then turns on a light according to a random schedule of known probability associated with each stimulus. Given, say, a probability of .8 associated with the left-hand stimulus, it is posited that the subject, in responding with the left-hand 80% of the time—which he will eventually do—will learn kinesthetic and proprioceptive cues, with knowledge of results as reinforcement. Let us term this left-hand response R_a and the stimulus to which it is made S_a. The right-hand response (less frequent) will be termed R_b and the stimulus associated with it S_b.

Given the above training procedure, with the probability of S_a equal to .8 and the probability of S_b equal to .2, if the competing response position is correct, will result in S_a being a potent elicitor of R_a in subsequent situations. S_b, given this situation, will be a less potent elicitor, both in relation to R_b (with a probability of .2), and to R_a (with which it should compete). Note that this position does not assume that R_a will be made to S_b, but merely that the situation will set up a response competition which should have discriminable effects with respect to the TIR phenomenon, and which should vary according to whether S_a precedes or follows S_b in the double-stimulation paradigm. This procedure has never, to the writer's knowledge, been used to test the underlying assumptions of a competing-response position.

REFERENCES

Adams, J. A. Human tracking behavior. *Psychol. Bull.*, 1961, **58**, 55–79.

Adams, J. A. Test of the hypothesis of psychological refractory period. *J. exp. Psychol.*, 1962, **64**, 280–287.

ADAMS, J. A., & CHAMBERS, R. W. Response to simultaneous stimulation of two sense modalities. *J. exp. Psychol.*, 1962, 63, 198–206.

ALLAN, MARY D. Memorizing, recoding, and perceptual organization. *Brit. J. Psychol.*, 1961, 52, 25–30.

BARTLEY, S. H. *Vision: A study of its basis.* (Rev. ed.) New York: Hafner, 1963.

BERLYNE, D. E. Uncertainty and conflict: A point of contact between information theory and behavior theory concepts. *Psychol. Rev.*, 1957, 64, 329–339.

BERLYNE, D. E. *Conflict, arousal, and curiosity.* New York: McGraw-Hill, 1960.

BORGER, R. The refractory period and serial choice-reactions. *Quart. J. exp. Psychol.*, 1963, 15, 1–12.

BROADBENT, D. E. The role of auditory localization in attention and memory span. *J. exp. Psychol.*, 1954, 47, 191–196.

BROADBENT, D. E. Successive responses to simultaneous stimuli. *Quart. J. exp. Psychol.*, 1956, 8, 145–152.

BROADBENT, D. E. Immediate memory and simultaneous stimuli. *Quart. J. exp. Psychol.*, 1957, 9, 1–11. (a)

BROADBENT, D. E. A mechanical model for human attention and immediate memory. *Psychol. Rev.*, 1957, 64, 205–215. (b)

BROADBENT, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

BRYDEN, M. P. Order of report in dichotic listening. *Canad. J. Psychol.*, 1962, 16, 291–299.

CREAMER, L. R. Event uncertainty, psychological refractory period, and human data processing. *J. exp. Psychol.*, 1963, 66, 187–194.

DAVIS, R. The limits of the "psychological refractory period." *Quart. J. exp. Psychol.*, 1956, 8, 24–38.

DAVIS, R. The human operator as a single channel information system. *Quart. J. exp. Psychol.*, 1957, 9, 119–129.

DAVIS, R. The role of "attention" in the psychological refractory period. *Quart. J. exp. Psychol.*, 1959, 11, 211–220.

DAVIS, R. Choice reaction times and the theory of intermittency in human performance. *Quart. J. exp. Psychol.*, 1962, 14, 157–166.

ELITHORN, A., & LAWRENCE, C. Central inhibition: Some refractory observations. *Quart. J. exp. Psychol.*, 1955, 7, 116–127.

GOTTSDANKER, R., BROADBENT, L., & VAN SANT, C. Reaction time to single and first signals. *J. exp. Psychol.*, 1963, 66, 163–167.

HALLIDAY, A. M., KERR, M., & ELITHORN, A. Grouping of stimuli and apparent exceptions to the psychological refractory period. *Quart. J. exp. Psychol.*, 1960, 12, 72–89.

HELSON, H. Current trends and issues in adaptation-level theory. *Amer. Psychologist*, 1964, 19, 26–38.

HELSON, H., & STEGER, J. A. On the inhibitory effects of a second stimulus following the primary stimulus to react. *J. exp. Psychol.*, 1962, 64, 201–205.

HYMAN, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, 45, 188–196.

KLEMMER, E. T. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, 51, 179–184.

MORAY, N. Attention in dichotic listening: Affective cues and the influence of instructions. *Quart. J. exp. Psychol.*, 1959, 11, 56–60.

MORIN, R. E., & FORRIN, B. Response equivocation and reaction time. *J. exp. Psychol.*, 1963, 66, 30–36.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

POULTON, E. C. Perceptual anticipation and reaction time. *Quart. J. exp. Psychol.*, 1950, 2, 99–112.

RESTLE, F. *The psychology of judgment and choice.* New York: Wiley, 1961.

REYNOLDS, D. *An investigation into some perceptual correlates of prejudice.* Unpublished master's thesis, Michigan State University, 1962.

REYNOLDS, D., & TOCH, H. H. Perceptual correlates of prejudice: Stereoscopic constancy using biracial stereograms. *J. soc. Psychol.*, in press.

TELFORD, C. W. The refractory phase of voluntary responses. *J. exp. Psychol.*, 1931, 14, 1–35.

TREISMAN, ANNE. Contextual cues in selective listening. *Quart. J. exp. Psychol.*, 1960, 12, 242–248.

VINCE, MARGARET A. The intermittency of control movements and the psychological refractory period. *Brit. J. Psychol.*, 1948, 38, 149–157.

WELFORD, A. T. The psychological refractory period and the timing of high speed performance: A review and a theory. *Brit. J. Psychol.*, 1952, 43, 2–19.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1954.

(Received February 13, 1964)

Psychological Bulletin
1964, Vol. 62, No. 5, 348–350

THE STRUCTURAL BASIS OF TIMING [1]

STUART J. DIMOND

Trinity College, Dublin, Ireland

An attempt is made to find the structural basis of timing. Although the physiological evidence allows no precise statement, there is some evidence for the involvement of anatomical structures. This evidence is reviewed, and a theory is proposed as to the function of these areas in timing.

The stream of sensory impulses gathered from the environment is distributed in time. It is supposed that the duration of stimuli and the intervals between them are compared with an internal standard. Such a standard could be represented by the steady functioning of some mechanism of the body. This role has been attributed to many mechanisms which function in a rhythmic manner, from the heart to the metabolic processes of the cells themselves. François (1927) found a relationship between the rate at which a key was tapped and the subject's body temperature during normal diurnal fluctuation and in the process of diathermy. Hoagland (1934) found that the rate at which his wife counted depended on her body temperature (when she suffered from influenza). Werboff (1957) established that time judgment was related to the amount of alpha activity measured by EEG studies. Stevens (1902) discovered that the fluctuations in time judgment coincided 50% of the time with changes in blood volume as measured by the finger plethysmograph. Schaefer and Gilliland (1938) on the other hand found no relationship of time estimation to: pulse rate, heart work, breathing rate, breathing work, and blood pressure changes.

[1] My thanks are due to Stephen Griew and Derek Forrest for their help in preparing this manuscript.

Gardner (1935) found some relationship between time estimation and disorder of the thyroid. It is well known that thyroxin influences the amount of heat produced within the body, by oxidizing substances within the cells themselves. Changes in body temperature could then cause changes in time estimation, for a relationship which has already been described exists between body temperature and time estimation (François, 1927; Hoagland, 1934). The increased production of thyroid hormone leads to increased enzymatic activity, which causes the strength of the heart beat to be increased, so that not only is thyroid disturbance accompanied by change in time estimation but also by changes in blood pressure. It would be expected therefore that changes in time estimation and blood pressure are related. This relationship was demonstrated in a previously mentioned study by Stevens (1902), but not by Schaefer and Gilliland (1938) who produced entirely negative results. All that can be said so far on the basis of this evidence is that the operation of the timing system is susceptible to those factors which influence the basic metabolism of cells.

Although the physiological evidence allows no more precise statement than this, there is some evidence for an involvement of anatomical structures. Remy (1942) reports the induction of the Korsakoff syndrome with con-

sequent temporal disturbance following lesions of the mammillary bodies; Tsiminakis (1931) found however that there was considerable variation in the appearance of the sydrome following such lesions. Temporal disturbance has been observed by Ehrenwald (1931) following lesion to the parietal lobe. Partridge (1950) and Hyde and Wood (1949) report similar disturbance following prefrontal lobotomy. Spiegel and Wycis (1949) found transitory temporal confusion following dorsomedial thalamotomy. They further investigated this phenomenon in thalamotomy and found considerable temporal confusion (Spiegel, Wycis, Orchinik, & Freed, 1955).

Although Tsiminakis (1931) has doubted that lesion of the mammillary bodies is consistently followed by temporal confusion, nevertheless, other investigators, namely, Harrison (1940) and Remy (1942), have found this to be the case. From the mammillary bodies, heavy projection fibers pass to the thalamic nuclei in the form of the mammillo-thalamic tract (Harrison, 1940). The relay functions of the thalamus are well known: the great sensory tracts and the somesthetic sensory pathways terminate in the posteroventral nucleus of the thalamus from which fibers pass to the cerebral cortex. Temporal confusion following lesion of the thalamus has been established (Spiegel & Wycis, 1949). The thalamus could act as a relay station to transmit, perhaps from the mammillary bodies, the neural impulses which from the basis of the internal time standard. Tumors involving the thalamus are often associated with diminution of amplitude or absence of alpha activity (Kirstein, 1953) and it is notable that: time judgment is also related to the amount of alpha activity during a particular interval (Werboff, 1957), and the introduction of a warning signal, while markedly reducing reaction time, also acts to block alpha activity (Lansing, Schwartz, & Lindsley, 1959). The thalamus performs the important function of relaying impulses from the great sensory tracts and from the spinothalamic tract to the cortex, so that lesions of the thalamus could also disrupt the process allowing information about time from the environment to reach the cortex.

Temporal confusion has been reported following lesion of the prefrontal lobes (Hyde & Wood, 1949; Partridge, 1950). The prefrontal lobes are noted for their planning function. Surgical removal impairs ambition and the ability to plan for the future. In rats (Carpenter, 1952; Epstein & Morgan, 1943; Maher, 1955; Steller, Morgan, & Yarosh, 1942) and in monkeys (Jacobsen, 1939), disorders of serial learning are especially prominent after frontal ablations. Animals with such lesions tend to make many more anticipatory errors than normal animals, or animals with more posterior lesions.

It is supposed therefore that the prefrontal areas of the cortex perform the fine acts of discrimination and judgment in the timing of events, and that the comparison of information about time reaching the cortex from the environment, with the internal standard, is made at this point.

REFERENCES

Carpenter, J. A. Anticipatory behavior in the rat following frontal lesions. *J. comp. physiol. Psychol.*, 1952, **45**, 413–418.

Ehrenwald, H. Störung der Zeitauffassung, der räumlichen Orientierung, des Zeichnens und des Rechnens bei einem Hirnverletzten. *Z. Neurol. Psychiat.*, 1931, **132**, 518–569.

Epstein, M. A., & Morgan, C. T. Cortical localization of symbolic processes in the rat: III. Impairment of anticipatory functions in prefrontal lobectomy in rats. *J. exp. Psychol.*, 1943, **32**, 453–463.

FRANÇOIS, M. Contribution a l'étude du sens du temps: La température interne, comme facteur de variation de l'appréciation subjective des durées. *Année psychol.*, 1927, 28, 186–204.

GARDNER, W. A. The influence of the thyroid gland on the consciousness of time. *Amer. J. Psychol.*, 1935, 47, 698–701.

HARRISON, F. The hypothalamus and sleep. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1940, **20**, 635–656.

HOAGLAND, H. The physiological control of judgments of duration. *Amer. J. Physiol.*, 1934, **109**, 54.

HYDE, R. W., & WOOD, A. C. Occupational therapy for lobotomy patients. *Occup. Ther. Rehabilit.*, 1949, 28, 109–124.

JACOBSEN, C. F. The effects of extirpation on the higher brain processes. *Physiol. Rev.*, 1939, **19**, 303–322.

KIRSTEIN, L. The occurrence of sharp waves, spikes and fast activity in supratentorial tumours. *EEG clin. Neurophysiol.*, 1953, **5**, 33–40.

LANSING, R. W., SCHWARTZ, E., & LINDSLEY, D. B. Reaction time and EEG activation under alerted and nonalerted conditions. *J. exp. Psychol.*, 1959, **58**, 1–7.

MAHER, B. A. Anticipatory and perseverative errors following frontal lesions in the rat. *J. comp. physiol. Psychol.*, 1955, 48, 102–105.

PARTRIDGE, M. *Prefrontal leucotomy*. Oxford, England: Blackwell, 1950.

REMY, M. Contribution a l'étude de la maladie de Korsakow: Étude anatomoclinique. *Mschr. Psychiat. Neurol.*, 1942, 106, 128–280.

SCHAEFER, V. G., & GILLILAND, A. R. The relation of time estimation to certain physiological changes. *J. exp. Psychol.*, 1938, 23, 545–552.

SPIEGEL, E. A., & WYCIS, H. T. Physiological and psychological results of thalamotomy. *Proc. Roy. Soc. Med.*, 1949, 42(Suppl.), 84–93.

SPIEGEL, E. A., WYCIS, H. T., ORCHINIK, C. W., & FREED, H. The thalamus and temporal orientation. *Science*, 1955, **121**, 771–772.

STELLAR, E., MORGAN, C. T., & YAROSH, M. Cortical localization of symbolic processes in the rat. *J. comp. Psychol.*, 1942, 34, 107–126.

STEVENS, H. C. The relation of fluctuation of judgement in the estimation of time intervals to vaso-motor waves. *Amer. J. Psychol.*, 1902, **13**, 1–27.

TSIMINAKIS, Y. Beitrag zur Pathologie der alkoholischen Erkrankungen des Zentral-Nervensystems. *Arb. Neurol. Inst. U. Wein*, 1931, 33, 24–62.

WERBOFF, J. The relationship between electroencephalographic activity and the estimation of short temporal intervals. Unpublished doctoral dissertation, University of Washington, 1957.

(Received January 28, 1964)

Psychological Bulletin
1964, Vol. 62, No. 5, 351–357

POSITIVE REINFORCEMENT BY LIGHT:

COMMENTS ON LOCKARD'S ARTICLE

C. C. KIERNAN[1]

Birkbeck College, London University

Lockard's (1963) review of studies exploring the conditions controlling light-onset and light-offset reinforcement and theories of light-reinforced bar pressing is discussed. The evidence is reexamined in terms of temporal and response variables and it is shown, contrary to Lockard's conclusions, that the data can then be seen as relatively systematic. Lockard's discussion of theories is criticized on the grounds that in the case of the facilitation hypothesis, the stimulus-change hypothesis, and the discrepancy hypothesis, the position discussed is much more powerful than Lockard suggests. Evidence contrary to Lockard's "preference theory" is cited in the course of the discussion.

This paper draws attention to conditions controlling the reinforcing effects of light onset and light offset in rodents which were not brought out by Lockard (1963, pp. 511–516) in his review of this field. They affect both his criticisms of other theories and the scope of the "preference theory" which he himself advances.

In his section on Demonstrations of Positive Reinforcement by Light, Lockard discusses several classes of variables affecting reinforcement and several theories. Space prevents discussion of all relevant points, and consequently only those sections in which major criticisms arise will be discussed. The sequence of Lockard's subsections will be followed.

Deprivation of Light. Two clear effects are apparent from the literature to which Lockard gives little weight. First, Premack and Collier (1962, Experiment III) demonstrated that with session effect (intersession interval) carefully controlled "the longer the rat occupies a given environment prior to being placed in test, the more light contingent bar presses it subsequently makes [p. 16]." Second, this effect was demonstrated with both light- and dark-maintained rats; that is, light deprivation as such did not appear as a significant variable.

Evidence contrary to the first finding (Robinson, 1959–60), cited by Lockard, was derived from a 20-inch tilt box. It seems possible that the radical difference in apparatus may render the two sets of data noncomparable at this stage.

Food and Water Deprivation. Here Lockard suggests that "a 'deprivation' effect may occur but shows no impressive consistency across studies." Considerable consistency can be introduced if the studies are divided on the basis of those including operant pretests before light-onset trials (Davis, 1958; Kling, Horowitz, & Delhagen, 1956; Premack & Collier, 1962; Segal, 1959), three of which (Davis, Premack, & Collier, & Segal) showed significant deprivation effects, as against the studies of Forgays and Levin (1958) and Wilson (1962), neither of which employed pretests and neither of which showed the effects of deprivation. The results of Kling, Horowitz, and Delhagen (1956) were derived from small groups, and

[1] The author is indebted to Arthur Summerfield for his kind assistance during the preparation of this paper.

Kling is reported by Davis (1958) as having found significant deprivation effects, but the conditions of the second experiment are not stated. In the remaining study Hurwitz and De (1958) did not employ pretests, or a satiated group, and used lengths of deprivation not used in other experiments. For the latter reason their results do not conflict with those of other studies, as is implied by Lockard, but suggest only that the function is probably a complex one.

A serious defect in the above analysis is that those studies showing negative results included neutral levers, or ran deprived rats under operant conditions, whereas the studies showing positive effects did not include these controls. It is suggested, however, that the evidence both here and elsewhere is sufficiently strong to warrant consideration of the pretest variable, and that the problems within this field can be resolved empirically and do not necessitate the establishment of conventions.

Effects across Time. In discussing these effects Lockard does not mention the clearly established effects of intersession interval, exhaustively investigated by Premack and Collier (1962) whose results demonstrate that LCBP is an increasing function of time between testing sessions. Further, he dismisses data relating to the adaptation-time hypothesis and the effect of pretests without adequate consideration. Appel and Hurwitz (1959) predict a fall from Day 1 to Day 2 in LCBP following unconditioned operant pretests, familiarization. Of the studies examined 15 support this prediction, although in several cases the decline is not indicated as significant (Appel & Hurwitz, 1959; Barry & Symmes, 1963; Davis, 1958; Ehrlich, 1961; Forgays & Levin, 1959, Group III; Girdner, 1953; Hurwitz, 1956; Hurwitz & Appel, 1959; Kling, Horowitz, & Delhagen, 1956; Leaton, Symmes, & Barry, 1963; Marx, Henderson, & Roberts, 1955; Premack & Collier, 1962; Roberts, Marx, & Collier, 1958; Robinson, 1959–60; Stewart, 1959). Two studies do not support the hypothesis. In the first, Forgays and Levin (1959, Groups I and II), the complexity of the reinforcement contingency (a 5-second light onset on depression of one of two levers) may be responsible since the same experiment shows the effect in another group (III) where the contingency was simpler.

Premack and Collier (1962) report two discrepant results. In Experiment IIB of their series a group tested every 48 hours showed an insignificant rise in LCBP over days. Since long intersession intervals are known to raise LCBP, and other studies have used shorter intervals (24 hours), this variable is probably responsible for the result; although the finding may still require the adaptation-time hypothesis to be modified. In the other discrepant instance, Experiment IID, dark-maintained rats showed a rise in LCBP from Day 1 to Day 2 of testing. It is possible that this result is partly a function of the low response rate shown by these animals.

The prediction implied by Appel and Hurwitz (1959), that when adaptation trials are not included in the design of an experiment a rise from Day 1 to subsequent days will be observed, is borne out by five of the eight relevant studies (Forgays & Levin, 1958, 1961—Distributed group; Hurwitz, 1956; Leaton, Symmes, & Barry, 1963; Levin & Forgays, 1959). Of the other three studies, Forgays and Levin (1961, Massed group) showed a nonsignificant decrease over blocks of trials. This result is probably accounted for by the fact that animals were run on a 1-hour intersession interval. However, both Henderson (1957) and Roberts (1962) show a de-

cline with 24-hour intersession intervals. Both results are difficult to accommodate, in neither case is there any clear factor which can account for the differences between these and other studies.

Theories of Light-Reinforced Bar Pressing

Facilitation Effects. Lockard (1963) claims that "evidence against the facilitation hypothesis is convincing [p. 515]," but does not discuss most of the evidence which supports it. The hypothesis suggests that increased activity and hence increased LCBP is brought about by increased stimulation (Leaton, Symmes, & Barry, 1963; Lubow & Tighe, 1957; Nash & Crowder, 1960; Wilson, 1962). Lockard suggests that in acquisition the hypothesis is ruled out by the results of Kling, Horowitz, and Delhagen (1956) who tried to control for facilitation effects by using a yoked-control technique. However Nash and Crowder (1960) utilizing a yoked-control technique showed that, although differences between experimental (light onset) and yoked controls were significant ($p < .001$) in an onset session, these differences disappeared under extinction ($p = .50$), suggesting that acquisition had not occurred. This experiment, which Lockard states used small groups, employed 29 experimental animals and 14 yoked controls. This result suggests that an adequate test for reinforcement should rely on extinction data.

Lockard cites extinction data as a further line of evidence against the facilitation hypothesis, but does not discuss the negative results within this area. The data suggest that extinction effects will be shown when between 10- and 20-response light-onset pairings are permitted (Barnes & Baron, 1961, Experiment II; Crowder, 1961, Experiments I, II, III; Crowder, Morris, Dyer, & Robinson, 1961; Crowder, Wilkes, & Crowder, 1960, Experiment I) or when a large number are permitted by including several onset sessions (Forgays & Levin, 1958; Hurwitz, 1956; Wilson, 1962). A third group of studies falling midway between these two on the variable of number of onset trials shows no significant extinction effects (Barnes & Baron, 1961, Experiment II; Crowder, Wilkes, & Crowder, 1960, Experiment II; Nash & Crowder, 1960). Two studies do not fit this scheme (Barnes, Kish, & Wood, 1959; Kish, 1955). In both of these, more unconditioned operant pretests were included than in other designs. The results in this field taken together suggest a probable complex relationship between number of response onset pairings and responding in extinction with a possible interaction with pretest responding. This relationship is not predicted by the facilitation hypothesis or by Lockard's preference theory.

The final evidence which Lockard brings to bear on the facilitation hypothesis is that of Forgays and Levin (1959) who showed discrimination and reversal learning under light-onset contingency. Lockard suggests that the facilitation hypothesis "could not handle a discrimination." However it may be argued that when stimulation occurs following LCBP the animal will be touching, or will just have touched, the onset lever. Hence stimulation will facilitate responses on, or in the region of, the onset lever. This position is backed by the results of Crowder, Wilkes, and Crowder (1960) who demonstrated "discrimination" in training sessions but showed no discrimination between levers under extinction conditions.

The demonstration of reversal learning does create difficulties for the facilitation hypothesis which would predict an immediate shift to the new onset lever rather than the gradual shift observed.

Two further experiments, not cited by Lockard, demonstrate probable facilitation effects and suggest problems for any reinforcement theory of LCBP. Thomas, Appel, and Hurwitz (1958) show a significant influence of lever size on responding in the light-onset situation; Forgays and Levin (1959) show a significant influence of number of levers (two against one).

While it is clear that facilitation effects are not capable of explaining a large body of data, it is nonetheless suggested that their influence is sufficiently powerful to warrant close consideration, especially in relation to acquisition data. No writer has suggested that the facilitation hypothesis can explain all the effects noted in the LCBP situation. Failure to consider these effects represents a major defect both at the experimental and theoretical level in this field.

Stimulus-Change Hypothesis. In this section Lockard notes that, in general, light offset has not been found to be reinforcing; and on the basis of this claim, and the weak evidence on effects of stimulus variation in rodent studies, he implies that the hypothesis does not provide a valuable approach to the data.

Of the nine relevant experiments to date on the topic of light offset, four have shown the effect. In all positive cases unconditioned operant pretests were included in the design (Barry & Symmes, 1963; Leaton, Symmes, & Barry, 1963; Roberts, Marx, & Collier, 1958; Stewart, 1959). In the case of those experiments not showing the effect, only Robinson included pretests (Barnes, Kish, & Wood, 1959; Hurwitz, 1956; Robinson, 1957, 1959; Wilson, 1962). Leaton, Symmes, and Barry (1963), in a parametric study, demonstrate that number of pretests is a crucial variable in revealing the light-offset effect. In view of this and the other evidence it would seem reasonable to suppose that Robinson's results are a function of one or more confounding variables, of which pretest responding is one probable candidate.

It would seem apposite to note in view of this evidence, and in view of the fact that Roberts, Marx, and Collier (1958) demonstrated that light- as well as dark-reared rats would turn off a light, Lockard's conclusion that the light-offset aspect of the Roberts, Marx, and Collier (1958) study could be classed with the light-aversion data is not justified (Lockard, 1963, p. 515). This constitutes an objection to Lockard's preference theory, which is backed by evidence of Stewart (1959). In the offset situation she used lights of .01 and 8.50 foot-candles, with light-maintained rats, and showed a significant effect.

The evidence as a whole therefore is against Lockard's (1963) conclusion that "the stimulus-change position is rendered almost untenable by the light offset studies [p. 516]." Furthermore the evidence of Premack and Collier (1962), here in the section Deprivation of Light, suggests an interpretation in terms of stimulus change.

Discrepancy Hypothesis. In his last discussion in this section, Lockard (1963) dismisses the discrepancy hypothesis as a contender. Two points arise: First, Lubow and Tighe (1957, p. 594) in fact found support for this hypothesis in that increases in intensity of stimulation were found to raise activity level, whereas Lockard suggests that they did not. Modified by these results the hypothesis predicts the usual finding that light offset is less reinforcing than light onset (Robinson, 1959; Stewart, 1959). Second, the hypothesis, in predicting on the basis of the "discrepancy between present stimulation and past stimulation" to changes in activity level (Lubow & Tighe, 1957), can

predict base level bar-pressing rates as a function of the discrepancy between maintenance and static test illumination. Response rates in LCBP would be seen as a function of the interaction of these base rates, and response rate dependent on the direction of change in illumination contingent on response. A more sophisticated form of the discrepancy hypothesis has been applied in this field (Bevan & Adamson, 1962), but is not discussed by Lockard. This hypothesis accurately predicts results not predicted by Lockard (Hurwitz, 1960), as well as the results of Roberts, Marx, and Collier (1958).

A conspicuous omission from the hypotheses considered by Lockard is the suggestion that light can be classed with reinforcers such as food and water. This view has stimulated several workers whose orientation warrants consideration, for example, Barnes and Baron (1961), Forgays and Levin (1959, 1961), Kish and Baron (1962), and Stewart and Hurwitz (1958). Further omissions include Berlyne's (1960) theory and Bindra's (1959) formulation as they apply in this field.

Conclusions

The treatment of light-onset and light-offset studies by Lockard (1963, pp. 511–516) has been discussed. He implied that the results were unsystematic (p. 514): it is suggested that in nearly every case these results can be interpreted in terms of the effects of temporal factors (or factors concerning the number of responses emitted). Unconditioned operant pretests, or adaptation time, influenced results in studies on deprivation of stimulus change, food and water deprivation, the form of the intersession performance curve under LCBP, and the occurrence of light-offset reinforcement effects. In the case of extinction data similar factors were evidenced here concerning the number of responses under the light-onset condition, with a possible interaction with adaptation time. Intersession interval was noted as showing a systematic effect on LCBP. These effects are in addition to those of maintenance illumination. It is suggested that discrepancies among results can be resolved without recourse to "apparatus standardization and procedural conventions [Lockard, 1963, p. 514]."

In considering theories covered by Lockard it was found that the facilitation hypothesis, the stimulus-change hypothesis, and the discrepancy hypothesis were all more strongly supported than Lockard had suggested.

Much of the evidence described in this paper is beyond the scope of Lockard's (1963) preference theory in its present form. Other evidence not referred to in this paper creates further difficulties for the theory, for example, Crowder and Crowder (1961). Alternative theories, such as Helson's (1959) adaptation-level theory, the Bevan-Adamson (1962) theory of reinforcement, or Bindra's (1959) novelty-reactions theory, seem to offer the opportunity to integrate the work of Lockard and others in a relatively clear and detailed fashion, rather than dismissing much of it as merely capricious.

REFERENCES

APPEL, J. B., & HURWITZ, H. M. B. Studies in light-reinforced behaviour: IV. Effects of apparatus familiarization. *Psychol. Rep.*, 1959, **5**, 355–356.

BARNES, G. W., & BARON, A. The effects of sensory reinforcement on extinction behavior. *J. comp. physiol. Psychol.*, 1961, **54**, 461–465.

BARNES, G. W., KISH, G. B., & WOOD, W. O. The effect of light intensity when onset and termination of illumination is used as reinforcing stimulus. *Psychol. Rec.*, 1959, **9**, 53–60.

BARRY, H., & SYMMES, D. Reinforcing effects of illumination change in different phases of

the rat's diurnal cycle. *J. comp. physiol. Psychol.*, 1963, **56**, 117–119.

Berlyne, D. E. *Conflict, arousal and curiosity.* New York: McGraw-Hill, 1960.

Bevan, W., & Adamson, R. Internal referents and the concept of reinforcement. In N. F. Washburne (Ed.), *Decisions, values and groups.* Vol. 2. New York: Pergamon Press, 1962. Pp. 453–472.

Bindra, D. Stimulus change, reactions to novelty, and response decrement. *Psychol. Rev.*, 1959, **66**, 96–103.

Crowder, W. F. Weak-light reinforcement and flash distribution. *J. Psychol.*, 1961, **51**, 439–442.

Crowder, W. F., & Crowder, T. H. Duration of weak-light reinforcement. *Psychol. Rep.*, 1961, **8**, 130.

Crowder, W. F., Morris, J. B., Dyer, W. R., Robinson, J. V. Resistance to extinction and number of weak-light reinforcements. *J. Psychol.*, 1961, **51**, 361–364.

Crowder, W. F., Wilkes, W. P., & Crowder, T. H. Weak-light reinforcement with and without control for response facilitation. *J. Psychol.*, 1960, **49**, 181–184.

Davis, J. D. The reinforcing effect of weak-light onset as a function of food deprivation. *J. comp. physiol. Psychol.*, 1958, **51**, 496–498.

Ehrlich, A. The effects of past experience on the rat's response to novelty. *Canad. J. Psychol.*, 1961, **15**, 15–19.

Forgays, D. G., & Levin, H. Learning as a function of change of sensory stimulation in food-deprived and food-satiated rats. *J. comp. physiol. Psychol.*, 1958, **51**, 50–54.

Forgays, D. G., & Levin, H. Discrimination and reversal learning as a function of change in sensory stimulation. *J. comp. physiol. Psychol.*, 1959, **52**, 191–194.

Forgays, D. G., & Levin, H. Learning as a function of change of sensory stimulation: Distributed vs. massed trials. *J. comp. physiol. Psychol.*, 1961, **54**, 59–62.

Girdner, J. B. An experimental analysis of the behavioral effects of a perceptual consequence unrelated to organic drive states. Unpublished doctoral dissertation, Duke University, 1953.

Helson, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. *Sensory, perceptual, and physiological formulations.* New York: McGraw-Hill, 1959. Pp. 565–621.

Henderson, R. L. Stimulus-intensity dynamism and secondary reinforcement. *J. comp. physiol. Psychol.*, 1957, **50**, 339–344.

Hurwitz, H. M. B. Conditioned responses in rats reinforced by light. *Brit. J. anim. Behav.*, 1956, **4**, 31–33.

Hurwitz, H. M. B. The effect of illumination conditions on the effectiveness of light-onset as a reinforcer: A test of the Bevan-Adamson reinforcement theory. *Brit. J. Psychol.*, 1960, **51**, 341–346.

Hurwitz, H. M. B., & Appel, J. B. Light-onset reinforcement as a function of the light-dark maintenance schedule for the hooded rat. *J. comp. physiol. Psychol.*, 1959, **52**, 710–712.

Hurwitz, H. M. B., & De, S. C. Studies in light reinforced behavior: II. Effect of food deprivation and stress. *Psychol. Rep.*, 1958, **4**, 71–77.

Kish, G. B. Learning when the onset of illumination is used as reinforcing stimulus. *J. comp. physiol. Psychol.*, 1955, **48**, 261–264.

Kish, G. B., & Baron, A. Satiation of sensory reinforcement. *J. comp. physiol. Psychol.*, 1962, **55**, 1007–1010.

Kling, J. W., Horowitz, L., & Delhagen, J. E. Light as a positive reinforcer for rat responding. *Psychol. Rep.*, 1956, 2, 337–340.

Leaton, R. N., Symmes, D., & Barry, H. Familiarization with the test apparatus as a factor in the reinforcing effect of change in illumination. *J. Psychol.*, 1963, **55**, 145–151.

Levin, H., & Forgays, D. G. Learning as a function of sensory stimulation of various intensities. *J. comp. physiol. Psychol.*, 1959, **52**, 195–201.

Lockard, R. B. Some effects of light upon the behavior of rodents. *Psychol. Bull.*, 1963, **60**, 509–529.

Lubow, R. E., & Tighe, T. J. A test of the discrepancy hypothesis of motivation using intensity of visual stimulation. *J. comp. physiol. Psychol.*, 1957, **50**, 592–595.

Marx, M. H., Henderson, R. L., & Roberts, C. L. Positive reinforcement of the bar-pressing response by a light stimulus following dark operant pretests with no aftereffect. *J. comp. physiol. Psychol.*, 1955, **48**, 73–76.

Nash, F. A., & Crowder, W. F. A further study of weak-light reinforcement and response facilitation. *J. Psychol.*, 1960, **50**, 287–290.

Premack, D., & Collier, G. Analysis of nonreinforcement variables affecting response probability. *Psychol. Monogr.*, 1962, **76**(5, Whole No. 524).

Roberts, C. L. Stimulus and stimulus-change factors governing the free operant rate. *J. comp. physiol. Psychol.*, 1962, **55**, 375–380.

Roberts, C. L., Marx, M. H., & Collier, G. Light onset and light offset as reinforcers for the albino rat. *J. comp. physiol. Psychol.*, 1958, **51**, 575–579.

Robinson, J. S. Light as a reinforcer for bar pressing in rats as a function of adaptation illumination level and direction of light change. *Amer. Psychologist*, 1957, **12**, 411. (Abstract)

Robinson, J. S. Light onset and termination as reinforcers for rats living under normal light conditions. *Psychol. Rep.*, 1959, **5**, 793–796.

Robinson, J. S. Final technical report. Technical Report; October 14, 1959–August 25, 1960; Contract No. AF 49 (638)-744, United States Air Force.

Segal, E. F. Confirmation of a positive relation between deprivation and the number of responses emitted for light reinforcement. *J. exp. Anal. Behav.*, 1959, 2, 165–169.

Stewart, J. Effects of light intensity and reinforcement scheduling on the positive and negative properties of light stimulation. Unpublished doctoral dissertation, London University, 1959.

Stewart, J., & Hurwitz, H. M. B. Studies in light-reinforced behaviour: III. The effect of continuous, zero and fixed-ratio reinforcement. *Quart. J. exp. Psychol.*, 1958, **10**, 56–61.

Thomas, D. G., Appel, J. B., & Hurwitz, H. M. B. Studies in light reinforced behaviour: V. Effects of lever size, shift in lever size and lever position. *Psychol. Rep.*, 1958, **4**, 411–413.

Wilson, J. J. Photic reinforcement as a function of optimal level of stimulation. *Psychol. Rec.*, 1962, **12**, 17–23.

(Received March 13, 1964)

Psychological Bulletin
1964, Vol. 62, No. 5, 358–360

PREADAPTATION: PANACEA FOR PAST PUZZLES

REPLY TO KIERNAN

ROBERT B. LOCKARD
University of Washington

Kiernan has placed a weighty explanatory burden upon the preadaptation hypothesis. While it is true that some evidence favorable to the position can be assembled by study grouping, it remains to be seen if the preadaptation hypothesis actually has the explanatory power attributed to it.

Kiernan's comments (1964) constitute two messages, one diffuse and the other succinct. The diffuse message is that I gave little weight to or failed to discuss in detail a collection of phenomena and hypotheses within one of the six sections of my article (Lockard, 1963). In an initial review of a formerly undefined aggregate of briefly explored effects and newborn hypotheses of the most tentative kind, the decisions on matters of depth, emphasis, and extent of discussion are made by the reviewer. At a time when there exists extreme disagreement about the "real reason" for the reinforcing effects of light and what more thoroughly explored phenomena this effect may be related to, it would seem unusual to admonish the reviewer with a list of minor objections about the relative space allocated to some of the topics discussed. Furthermore, it is anachronous to base some of these objections upon evidence from articles appearing after the preparation of my manuscript (Barry & Symmes, 1963; Kish & Baron, 1962; Leaton, Symmes, & Barry, 1963; Roberts, 1962).

The succinct message which Kiernan's (1964) article repeatedly delivers is that sorting studies on the basis of preadaptation trials, or none, will significantly reduce the apparent inconsistency between experiments in the presently puzzling areas of:

> studies on deprivation of stimulus change, food and water deprivation, the form of the intersession performance curve under LCBP, and the occurrence of light offset . . . effects [p. 355].

This extended application of the Appel and Hurwitz (1959) hypothesis may, or may not, have the properties attributed to it. While a very few papers (e.g., Leaton et al., 1963) have clearly shown that dependent variables assume different values at different levels of the preadaptation variable, it would seem quite an inductive leap to presume such sweeping organizational powers for a multidimensional and continuous variable considered only in its dichotomous form. Too few LCBP-type studies have deliberately explored the effects of different levels of adaptation trials upon the phenomena Kiernan asserts are brought into harmony by consideration of this particular procedural variable and not some other. Furthermore, if the preadaptation variable were so potent, one would expect it to figure prominently in better understood reinforcement situations and not to be unique to light-controlled behavior. The preadaptation hypothesis already existed in the literature, and I fail to appreciate what contribution to science is made by using past studies as data and grouping them

in an R-R manner to support a hypothesis. The issue is properly approached with S-R procedures in the laboratory, not with words and arguments. My own efforts to resolve between-study discrepancies left the matter open, for I did not wish to artificially simplify a possible complex phenomenon, and I did not choose to gamble that a single procedural variable would harmonize the discord (Lockard, 1963):

> the general picture is that any given study may find "an effect"; yet the way the effect operates comes out differently in different studies. Quite possibly the effect operates one way at some level of some additional variable and another way at some other level. Since two studies in disagreement often differ in many ways—strain of animals, apparatus, light intensity, adaptation trials—it is impossible to account for the disagreement [p. 514].

Theories of Light-Reinforced Bar Pressing

Kiernan fails to distinguish sharply between theories purporting to explain why dim light-onset is reinforcing and theories asserting that a particular variable has some effect in the LCBP situation. When a rat is put into a Skinner box with a bar connected to a light, all formerly discovered behavioral effects accumulated in the journals should be demonstrable under proper conditions. These would include adaptation trials, facilitation effects, discrepancy from adaptation level, and all other treatments known to affect organisms. Thus a truly complete review would extend in all directions to the limits of behavioral knowledge. Even though my original manuscript fell short of this, the editor's request to shorten it by 20 pages required brevity in secondary issues. I judged as secondary a group of procedural variables likely to operate across reinforcers and not likely to be unique to light-controlled behavior. Kiernan's comments would indeed be apposite if the title of my article had been something like "The Effects of Nonnutritive Variables upon Light-Reinforced Behavior." Since the title and defined content were "Some Effects of Light upon the Behavior of Rodents," the article was principally concerned with the primary variables. It is by no means certain that the LCBP situation is the optimum context for a discussion of procedural effects operating perhaps more clearly when reinforcers other than light are used. Until the reinforcer in the LCBP context is isolated, it may be unenlightening and premature to expend retroactive energy to conceptually organize the effects of secondary variables. Light seems to be a somewhat impotent controller of behavior, at times dominated by a number of other variables as procedural variations and static variables shift the delicate balance among competing processes. Contrary to Kiernan's inference, it is the experiments, not the phenomena, which are capricious. As a single example, it is known that light-controlled behavior varies with light intensity. Yet if one goes through the literature paper by paper, it is possible to determine the actual intensity of illumination falling upon a subject in only about 10% of the papers. When a body of literature is in the primitive stage of disregarding accurate reporting of the levels of its primary variable, it would seem unlikely that separate competing weak effects would be partitioned through an exegesis of past research by historians. New, clever, and exacting experimental contributions of the future will reveal the underlying harmony in natural phenomena.

The core problem of why the onset of dim light is reinforcing prompted the offering of the preference hypothesis as a provisional explanation. It states that, if the preference value of the stimulus

situation produced by responding exceeds that prevailing as a consequence of not responding, response strength will exceed that of suitable controls. It is not a circular theory, for the preference values of stimuli are determined by preference-testing procedures before predictions about their reinforcing properties are made. There are still many problems: for example, the effects of numerous variables upon preference behavior are yet unknown. However, the preference hypothesis bears credibly upon explaining why dim ambient light is reinforcing when darkness is the alternative of not responding. It accounts for the usually negative light-offset findings, unites the LCBP studies with the older light-aversion research, takes into account light-history, predicts differences between pigmented and albino strains, and subsumes more phenomena under a single hypothesis than other theories. It may be wrong, and it is surely incomplete; and if it is not subsumed under even more encompassing formulations, behavioral science will have failed in its effort to formulate a hierarchy of laws.

REFERENCES

Appel, J. B., & Hurwitz, H. M. B. Studies in light-reinforced behaviour: IV. Effects of apparatus familiarization. *Psychol. Rep.*, 1959, **5**, 355–356.

Barry, H., & Symmes, D. Reinforcing effects of illumination change in different phases of the rat's diurnal cycle. *J. comp. physiol. Psychol.*, 1963, **56**, 117–119.

Kiernan, C. C. Positive reinforcement by light: Comments on Lockard's article. *Psychol. Bull.*, 1964, **62**, 351–357.

Kish, G. B., & Baron, A. Satiation of sensory reinforcement. *J. comp. physiol. Psychol.*, 1962, **55**, 1007–1010.

Leaton, R. N., Symmes, D., & Barry, H. Familiarization with the test apparatus as a factor in the reinforcing effect of change in illumination. *J. Psychol.*, 1963, **55**, 145–151.

Lockard, R. B. Some effects of light upon the behavior of rodents. *Psychol. Bull.*, 1963, **60**, 509–529.

Roberts, C. L. Stimulus and stimulus-change factors governing the free operant rate. *J. comp. physiol. Psychol.*, 1962, **55**, 375–380.

(Received April 14, 1964)

Vol. 62, No. 6

December 1964

Psychological Bulletin

ENVIRONMENTAL VARIATION IN STUDIES OF ORGANIZATIONAL BEHAVIOR

GARLIE A. FOREHAND and B. von HALLER GILMER

Carnegie Institute of Technology

The desirability of incorporating environmental variation into research designs has often been pointed out, but few procedures for doing so have been developed. Attempts to operationalize the concept of "organizational climate" in studies of organizational behavior provide a number of methods for assessing environmental variation, and yield data relevant to hypotheses regarding the interaction of persons and environments. Methods for observing climate variation include field studies, assessments of participant's perceptions, observations of objective indices, and experimental control of organizational variables. Conditions may affect behavior by determining stimuli, by restraining freedom of response, and by rewarding and punishing behavior. Illustrative of the organizational properties meriting further study are size, structure, systems complexity, leadership pattern, and goal directions. Selected bibliography of 104 titles indicates how studies of organizations make possible the variation of both person and climate variables.

The concept of environment has been a difficult one for psychologists to deal with empirically. The postulate that behavior is a function of the interaction of organism and environment is widely accepted and both its theoretical and practical implications have been explored (Barton, 1961; Brunswik, 1956; Cronbach, 1957; Murray, 1938). Effort has been devoted to discovery of relevant dimensions of personal variables and to the precise definition of experimental treatments, but there have been few attempts to develop multivariate definitions of environment, and fewer still to study behavior as a function of the simultaneous variation of personal and situational factors. In the past few years several psychologists have noted the lack of such studies and have begun to develop models for systematically incorporating environmental variation in research design (e.g., Abelson, 1962; Sells, 1963a, 1963b). Models relevant to this problem of general psychology are emerging from the rapidly developing and multidisciplinary study of behavior of individuals in organizations.

The study of behavior of persons in organizations presents both a need and an opportunity for environmental analysis. Analyses of organizational behavior have multiple origins. An example of the diversity of approaches may be termed the institutional-individual continuum. Extremes on the continuum are represented by "classical" descriptions of organizations that derive their descriptive terms from formal, legalistic specifications (summarized by Strother, 1963) and by contemporary conceptions whose descriptive terms refer to the behavior of interacting individuals (Guetzkow & Bowes, 1957). Any seg-

ment of the continuum would, of course, have value in accounting for some observed events, and a given approach will handle some observations better than others. Not surprisingly, analyses focusing upon one end of the continuum are beset by devils whose home is at the other. Formal models of the organization and its parts provide neat symbolic devices, but members of the organization often do not behave as the model says they should (March & Simon, 1959). On the other hand, attempts to predict what an individual will do on the basis of his own personal characteristics often lead to the reluctant conclusion that behavior depends in part on the situation (Tagiuri, 1961).

Evidence that approaches to organizational behavior are coming together is summarized by Leavitt (1962a), who sees the emergence of "organizational psychology," a new multidisciplinary research field that "views organizations as curiosities deserving of research; and is therefore as much concerned with the nature of systems and the nature of human decision-making as with any applied problems [p. 27]." A consequence of such a perspective is intensified interest in human behavior as it is conditioned by organizational properties, and in organizations as they are influenced by the behavior of their members.

Organizational psychology offers unique opportunity for the study of environmental variation. Observers of—and participants in—organizations have noted differences in organizational personality or climate (Dill, Hilton, & Reitman, 1962; Gilmer, 1961), and some have written narrative descriptions reminiscent of early descriptions of personality types (e.g., Gellerman, 1960). The development of formal specification and measurement of such variation is a task of as much relevance to general psychology as to the specific study of organizations. An organization has several properties that make it a particularly appropriate focus for studying environmental variation: (*a*) The organization represents a component of the environment of behavior known to be influential on certain subsets of behavior. (*b*) The organization is bounded, at least in a relative sense; a company or a department can be identified unambiguously. And (*c*) there exist sources of information about an organization—in the form of records, statistical summaries, perceptions of participants, organization charts and the like—that provide potential data for characterizing the personality of the organization.

What to call the desired conceptually integrated synthesis of organizational characteristics has been a problem for many researchers. The term organizational climate has been used by several of them, although the term means different things to different writers. We shall use the term in this paper to refer to the set of characteristics that describe an organization and that (*a*) distinguish the organization from other organizations, (*b*) are relatively enduring over time, and (*c*) influence the behavior of people in the organization. These defining properties were chosen in the effort to focus discussion upon features of organizational variation that are amenable to specification, measurement, and incorporation into empirical research.

Work on the problem of conceptualizing variation in climate has been done on many fronts. Relevant studies are found in the literature of psychology, sociology, administration, and education. This paper reviews a number of these efforts, but is in no sense a survey. The pertinent literature is too diffuse to make such an attempt profitable at present. Rather, we have searched for

converging conclusions from samples of diverse approaches to the problem and have concentrated on studies in which variation in both climate and individual behavior was considered. Among the questions to be discussed are: How can organizational climate be measured? What dimensions of climate are meaningful? What kinds of relationships between climate and behavior may be expected? What do such relationships imply concerning the operation of organizations? What directions do these concepts suggest for future research?

Measurement of Organizational Variation

The measurement problem confronting the student of organizational climate is similar in some respects to that of the psychologist studying individual behavior. The problem may be conceived as one of the constructing tests for organizations, and thus involves the systematic observation of the behavior of organizations. As in individual testing there are a variety of possible ways for making the required systematic observations. Approaches include: intensive observation in "field" situations, assessments of perceptions of organizations by organization members, observation of objective organization properties, and experimental variation of organizational properties.

Field Studies

Intensive observations of the actual, ongoing activities of organizations may provide a researcher with a sensitive feel for the organization's climate, and, depending upon the skill of the observer, a communicable model of the organization's functioning. The sources of information available to the naturalistic observer are almost unlimited—observations of conferences, interviews with participants, diaries kept by participants, departmental memos, records and correspondence, and the history of the organization, to name only a few—and a wealth of descriptive material has been obtained in this way. Much of this material comes from case studies of single organizations. While hypotheses may sometimes be generated from such studies, an understanding of the effect of climate requires studies that systematically examine variations of climate as they influence the behavior of participants. Two approaches to the observation of climate variation have been (*a*) examining behavior in contrasting organizations (comparative studies) and (*b*) studying the effects of changing conditions in a single organization (longitudinal studies).

Designs for comparative studies have reflected varying perspectives of investigators. Two polar types of design will serve to illustrate the approach: the experimenter may begin with organizational systems that differ observably in properties hypothetically relevant to individual behavior, and look for differences in actual behavior; or he may focus upon groups whose behavior is observed to be contrasting and look for organizational correlates of that variation. The former design is illustrated by Barnes' (1960) study of two engineering work groups. The groups were approximately equal in size and had similar duties, but one was characterized by a "closed" authority system—tight control, low member autonomy, few opportunities for interaction of members, etc.—while the other presented a contrasting "open system." Other climate dimensions that have been chosen to define contrasts in field studies are informal social relations in small work groups (Blau, 1954), democratic versus authoritarian management policies (Stanton, 1960), and interdependence

of employees in a work group (Vroom & Mann, 1960).

The second approach—investigation of organization correlates of observed differences in behavior—was taken by Dill (1958) in a study of two firms that differed markedly in the degree of personal autonomy of their members, as indicated by differences in perceived autonomy and in observed conference behaviors. The two firms were found to differ both with respect to task environment (inputs of information from external sources, e.g., stability and homogeneity of customers, suppliers, competitors, and regulatory groups) and with respect to internal restraints characteristic of the firms themselves (e.g., stress on formal rules and procedures, top management involvement in routine activities).

An example of a longitudinal study of a single organization is provided by Argyris (1958). From a model of the process by which the climate of an organization evolves from input variables (e.g., hiring process, formal policies, leadership style) Argyris generated and tested hypotheses about what would occur in a particular organization under particular conditions. The hypotheses concerned, for example, what would happen if an officer who is not the right type (i.e., the type congruent with the climate) is appointed, if employees are requested to communicate their true feelings, and if officers are asked to diagnose the organization's human problems.

The richness of information available from field studies carries with it some serious disadvantages, notably the practical expense, demands upon the skill and sensitivity of the observer, the impracticality of obtaining a "sample size" of more than two or three organizations and, perhaps the most serious disadvantage, the inherent subjectivity of the classifications. Each of the remaining methods discussed here attempts to overcome one or more of these disadvantages, at the expense of detailed impressions.

Perceptions of Participants

Theoretical conceptions of the relationship of organizational properties to individual behavior often emphasize the role of perception of organizational properties as intervening variables. For example, Likert's interaction-influence model assigns central importance to organizational characteristics as they are perceived by the individual. The causal variables (structure, objectives, supervisory practices, etc.) interact with personality to produce perceptions, and it is only through perceptions that the relationship between causal and end-result variables may be understood (Likert, 1961, pp. 196 ff.). This point of view suggests the measurement of climate indirectly via the perception of the individuals whose behavior is being studied.

Another argument for assessing organizational climate by means of participant's perceptions is that such perceptions are based upon experience that is both more extensive and more involved than that of an outside observer. The developers of the College Characteristic Index (CCI) (Pace, 1962; Pace & Stern, 1958) make this rationale explicit:

> In answering the CCI, students act as observers, saying what they believe is generally true or not true of their college. The items refer to a wide range of topics—rules and regulations, facilities, student-faculty relationships, classroom methods, extra-curricular activities, etc. The argument is that all these characteristics and events and practices, added together, constitute an educational press on the awareness of students. The aggregate awareness of students about their college environments constitutes a press in the sense of exerting a directive influence [Pace, 1962, p. 47].

The CCI is one of a few inventories designed to assess a range of dimensions of climate. It contains 30 scales based upon Murray's concept of environmental press. The scales are named in terms of Murray's classification of needs, and include press toward humanism, impulsion deliberation, and reflectiveness. Thistlethwaite (1959, 1960) rescored the CCI to define scales reflecting faculty behavior as distinguished from student behavior. Hemphill (1956) has developed another set of scales for measuring dimensions of group performance. These 14 scales, based on factor-analytic studies, include measures of control, stability, and intimacy.

Other investigators have developed perceptually based measures of specific constructs, for purposes of testing specific hypotheses. For example, Forehand and Guetzkow (1962; Forehand, 1963) obtained a measure of executives' perceptions of the relative "rule centeredness" and "group centeredness" of their organizations by having subjects report the solutions to administrative problems that they thought would be most typically adopted in their organizations. Lodahl and Porter (1961) assessed necessity for work-group cooperation by means of judgments by supervisory personnel familiar with the work. Pelz (1951), in a study of employee behavior as a function of supervisors influence, judged the degree of supervisor's "influence over the social environment in which his employees were functioning" from the supervisor's reports concerning his voice in decisions made by his own supervisor, lack of formal contact with his supervisor, and his salary.

Measures based on perception have advantages in research convenience, and may have theoretical meaning in their own right, but in such a measure characteristics of the individual and the organization are confounded. As Sells (1963a) has pointed out, analysis of the interaction between person and situation will require independent identification of the variation in each. For such analyses, it will be necessary to define organizational variation in terms of objectively observable measures, or to develop means of systematically manipulating organizational properties.

Objective Indices

There have been several attempts to examine objective properties of organizations, properties easily and reliably obtainable from records. Several studies, for example, have been concerned with organizational size as a variable (Baumgartel & Sobol, 1959; Talacchi, 1960; Thomas, 1959). In a study of factors related to accidents, Sherman, Kerr, and Kosinar (1957) examined a long list of organizational variables, including union representation, extent of employee participation in incentive and profit-sharing plans, extent of seasonal layoffs, and the nature of the plant vicinity.

This method affords the possibility of studying a wide variety of organizations (e.g., the Sherman, Kerr, and Kosnar study included 147 plants), with consequent advantages for studying the generality of the conclusions. The principal difficulties of the procedure are the same as those confronted by the psychologist studying individual personality—the variables that may be examined are too numerous and too specific to be readily interpreted. Studies that examine in isolation specific objective properties of an organization leave unanswered the questions of how the properties are related to one another and how they are related to useful constructs of organizational functioning.

Evan (1963) has outlined an approach that has considerable promise for giving meaning to objective indices. Rather than taking administratively de-

fined measures as given, he begins with an abstractly defined concept that has hypothetical relevance to behavior and suggests ways of constructing indices to reflect that concept. Specifically, Evan has worked with the concept of degree of hierarchical organization. He postulates that "different degrees of hierarchical organization have different consequences for total and partial social systems," and outlines several indices of the hierarchies of skills, rewards, and authority in organizations. Suggested indices of the authority hierarchy, for example, involve (*a*) ratio of higher level supervisors to foremen; (*b*) number of levels of authority from top management to workers; (*c*) ratio of administrative to production personnel; (*d*) the ratio of maximum to minimum time span of control—that is, the length of time an employee is authorized to make organization-committing decisions on his own initiative; (*e*) degree of decentralization of defined classes of decisions; and (*f*) the number of echelons for which "procedural due process of law" has been institutionalized. Such an approach to organization measurement has the valuable effect of rendering the criteria of construct validity, particularly the ideas of convergent and discriminant validity (Campbell & Fiske, 1959), applicable to organizational measurement.

Another approach, taking a cue from differential psychology, applies multivariate statistical techniques to the analysis of organizational properties. In one such study, five situational characteristics of 72 wholesale warehousing divisions of a single firm were examined. The variables—size of work force, city size, wage rate, unionization and percentage of male employees—were found to be considerably intercorrelated and defined a single factor in a factor analysis. The factor was interpreted as:

> degree of urbanization of the setting in which the division is located, low urbanization being represented by relatively small community population, few employees in the division, lower wages, absence of a union, and lower proportion of male employees [Katzell, Barrett, & Parker, 1961, p. 67].

Palmer (1961) analyzed measures of 21 organizational conditions (e.g., pension plan, recreation facilities, quality control methods) and 9 personnel behaviors (e.g., productiveness, lateness, turnover) in 188 manufacturing firms. The variables could be accounted for by eight orthogonal factors, of which five—denoted retirement welfare, size of work force, thrift benefits, insurance benefits, and theft versus discounts—were defined by both organizational and behavioral measures.

Experimental Manipulation of Climate

Two characteristics of the organizational climate have been assumed in the foregoing discussion: that it is multidimensional in nature, and that it is characteristic of an existing organization, built upon factors beyond the investigator's control. We have thus talked of discovering rather than creating climate, although some investigators have envisioned the eventual possibility of selecting climates for maximum benefit. As a research device, the idea of identifying relevant dimensions of climate and varying them systematically has appeal. This would permit a clear-cut test of interaction hypotheses, among other advantages.

Such a strategy, applying to the concept of climate as we have defined it, is far from ready to be put to work, but there are several lines of effort that provide relevant models. The most detailed and voluminous work on the experimental control of social variables has involved small groups. Golembiewski (1962a, 1962b) and Leavitt and Bass

(1964) have recently surveyed literature in this field with its relevance to organizations in mind; although, as the latter investigators point out, "extrapolations from small group data [to large organizations] have been free and easy." The group properties singled out by Leavitt and Bass as potentially relevant to organizations are group maturity, size, group composition (homogeneity, cohesion, mutual esteem), and geographical distance. Golembiewski's longer reviews could be interpreted to suggest some 15–20 variables that might be defined experimentally, although they often are defined in other ways in small group studies. Examples are structural integration, status congruency, task, and threat.

Group experiments most clearly relevant to question about climate have come from two traditions. One is the study of communication networks (Glanzer & Glaser, 1961). These experiments permit variation of both the permitted pattern of communication and of individual position in the pattern. Highly "realistic" situations can be developed in this way; in a sense the experimenter actually "creates" an organization (Guetzkow & Bowes, 1957).

Perhaps the most extensive examinations of effects of organizational properties have developed from the "human relations" tradition in management. In attempts to demonstrate the relative superiority or inferiority of democratic, participative, employee-centered leadership practices, ingenious methods of varying organizational properties have been devised. Perhaps the most relevant studies here are those that build on the model of early research on leadership style (Lewin, Lippitt, & White, 1939) and whose experimental conditions are designed to reflect realistic characteristics of complex organizations.

Some of these experiments have approached the problem of realism by simulating organizational activities. For example, Kidd and Christy (1961) studied effects of several supervisory practices (laissez faire, direct participant, and active monitor) on productivity in a simulated air-traffic control center. In a well-controlled study, Day and Hamblin (1961) constructed a factorial design involving combination of close versus general supervision and punitive versus nonpunitive supervision in a simulated assembly-line situation. The independent variables were operationally defined by controlling the number of instructions given by supervisors and the number of sarcastic, negative, status-deflating remarks made by the supervisors, and by careful training of supervisors in the playing of their roles.

Several classic studies conducted by the Institute for Social Research have attempted experimental control of organizational variables in actual ongoing organizations. Coch and French (1948) varied the manner in which work groups were involved in a change in jobs. Experimental groups participated in decisions regarding the need and nature of job changes, with varying degrees of totality of participation in the planning, while a control group was instructed to carry out plans developed by the production department. In a further variation, participative treatment was introduced into the control group at a later time, and the production of this group before and after the change was contrasted. Morse and Reimer (1956) varied conditions of supervision in an organization by training supervisors in one group of offices to delegate and decentralize, and those in a similar group of offices to exercise closer supervision over employees.

An approach that has promise lies somewhere between the manipulation of existing organizations and the creation

of laboratory group situations. It involves the simulation of complex organizational activity, varying factors that are more complex than those of the experimental situation and less complex than those of the real organization (Guetzkow, 1962). Simulation has been used most often to date for purposes other than environmental variation: to provide a controlled context for training (Cohen & Rhenman, 1961), a constant background for the observation of behavioral processes (Chapman, Kennedy, Newell, & Biel, 1959), or constant conditions for the evaluation of individual performance (Hemphill, Griffiths, & Fredericksen, 1962). Simulation, however, offers a useful opportunity to vary some aspects of an organization while holding others constant, and for observing the effects of such variation on behavior. Bass (1963), for example, has developed simulated manufacturing organizations in which complexity of line-staff organization, among other properties, can be varied. Bass studied such variations with respect to both groups' productive efficiency (sales, costs, profits) and individual reactions (satisfaction, communication, style, and interpersonal relations).

Effects of Climate

The literature would profit from surveys both of studies of dimensions of organizations (along the lines of the Golembiewski, 1962b, analysis of small group dimensions) and of studies of the effects of organizational variables (along the lines of the Thomas and Fink, 1963, review of the effects of group size). In this section and the next we have set for ourselves somewhat more modest objectives: an examination of some possible mechanisms by means of which climate might affect behavior, and of some classes of variables that will need to be considered in any attempt to define dimensions. This strategy stems in part from the difficulty of assembling—or even finding—the relevant literature. But it may also be argued that such an approach is a necessary component of any attempt to get order into this field. For example, one of the few general hypotheses concerning organizational variation that has been investigated with any degree of thoroughness might be stated as follows: An organization in which personnel policies are participative, democratic, unstructured, will differ from one whose practices are nonparticipative, authoritarian, or structured in that productivity, and employee satisfaction will be higher. There is evidence in support of this hypothesis, much of it summarized by Likert (1961). There is other evidence that the hypothesis may be true for satisfaction but not for productivity (Blau & Scott, 1962; Morse & Reimer, 1956), and that it may be true for some jobs or some parts of an organization, but not for others (Leavitt, 1962b). Uncertainty about whether the hypothesis is generally true, or if not, when it is true and when it is not, will have to prevail until certain questions are asked. First, what organizational properties differentiate a structured from a nonstructured environment? Barnes' (1960) open system, Morse and Reimer's (1956) decentralization of authority, and Day and Hamblin's (1961) general (as opposed to close) supervision are interpreted similarly but are defined by widely differing operational criteria. Can we assume that they have some organizational property in common to a degree sufficient to permit comparison of their effects? These questions concern dimensions of organizational variation and are considered further in the next section. The second question is: Why should we expect such an organizational property to have such an effect? With respect to the specific

hypothesis stated above, answers to the question seem to rest more often on ideological than on theoretical grounds. The question of the way in which the environment impinges on the individual to affect behavior needs to be asked if the hypothesis is to have significance beyond its application in a specific context.

Two kinds of influence may be distinguished. A particular organizational property may influence the behavior of all or almost all members of the organization. This kind of relationship will be termed "direct influence": The general hypothesis of direct influence is that aggregate behavior under one organizational condition will differ from that under another. The second kind of influence, which will be termed "interactive influence," exists when an organizational condition has a certain effect upon the behavior of some independent identifiable persons, but another effect, or no effect, on that of others.

A number of studies have examined the influence of particular organizational variables, but few attempts have been made, either before or after the data are in, to posit a mechanism for such influence. Empirical results, therefore, give few clues as to the nature of such mechanisms. The following classification of ways in which climate may affect behavior is an attempt to formulate and clarify hypotheses concerning such effects with the realization that available results will not fall neatly into the categories. Organizational climate may affect behavior by: defining the stimuli which confront the individual, placing constraints upon the freedom of choice of behavior, and/or rewarding and punishing behavior.

Definition of Stimuli

If behavior of organization members is determined in part by their perceptions of the organizational environment, as several observers (e.g., Likert, 1961) have postulated, then behavior will be influenced by variations in the objects or events available to be perceived, that is, in stimuli. One difficulty in examining such effects is that of specifying the meaning of stimulus in a complex situation. Arnoult (1963) has described a social stimulus, in response to Gibson's (1960) questions, as follows:

> (a) it may have motivational properties; (b) it is the *occasion* rather than the *cause* of a response . . . ; (c) it must frequently be defined in terms of the response it elicits; (d) it exists in the environment and not just at the receptors; (e) it can be a pattern or sequence of events; (f) its structure must often be inferred from the structure of the response which it elicits; and (g) it carries information about its source in the environment [pp. 19–20].

Thus defined stimulus variation is essentially equivalent to environmental variation as we have discussed it here. Other investigators have also approached the entire problem in terms of stimulus properties of the situation (Helson, Blake, & Mouton, 1958; Rosenbaum, 1956; Sells, 1963c). The emphasis of this section is on Property b of Arnoult's classification and here we are concerned with ways in which organizational variation brings about variation in the occasion of a response.

The physical properties of organizational situations are recognized more or less explicitly in systems research. Chapman et al. (1959), for example, outlined procedures used in the experimental evaluation of an air control system for presenting stimuli, controlling extraneous stimuli, and directing the attention of subjects. Stimulus elements of machine design have been a classical focus for research on man-machine systems (Gregg, 1961).

A second, and more complex, set of

occasions for response consists of the task-relevant information available to the individual. The likelihood that information is differentially distributed among members of an organization is suggested by Cyert and March's (1963, pp. 101 ff.) model of the organization as a communication system. One of the functions performed by operating procedures in such a system is the distribution and condensation of information that comes into the organization. The effect of social structure upon the transmission of this information is illustrated by Blau's (1954) study of interviewers in an employment agency. Competitiveness among work group members—which varied with "climate" conditions, such as security of employment, opportunity for development of a common professional orientation, and supervisory evaluation procedures—was manifested most clearly in the hoarding of information concerning job openings: members of the more competitive group resorted to various illicit means to withhold such information from their colleagues, in an effort to maximize their comparative record of job placements. Similar information-withholding practices have been observed by Argyris (1962). One source of behavioral effects of variations in information transmission is the likelihood that, regardless of the motivational and freedom-of-choice variables in a situation, some behaviors never occur because the stimuli that would elicit them are never presented. Another is suggested by evidence that one's expectations and evaluation of his own performance are influenced by information about the expectations and evaluation of coworkers (Neimark & Rosenberg, 1959; Rosenberg & Hall, 1958; Zander & Curtis, 1962).

A third subset of stimulus elements that depend upon climate variables includes the behavior of individuals and social interactions within the organization. Organizations may define certain patterns of behavior to be appropriate, as discussed below, and thus restrict the social stimuli available to organization members. Studies of interpersonal perception have indicated that social judgments are influenced by a number of properties of the perceived, including status, role, and visibility of the trait to be judged (Costello & Zalkind, 1963, pp. 46–47). All of these properties can be influenced by organizational climate; the visibility variable will serve as an illustration. Goffman (1959) suggests that the stimulus qualities of a person's behavior result from a more-or-less conscious role-taking strategy. An organizational rule of thumb that limits or defines appropriate role-taking behaviors (e.g., formality of relationships with subordinates) might, therefore, suppress or heighten the visibility of a particular behavioral cue.

There is, of course, much evidence that perceptions are influenced by abilities, values, and personality traits of the perceiver (Bruner, 1958; Costello & Zalkind, 1963, pp. 45–46; Taft, 1955) and by his organizational role (Dearborn & Simon, 1958). Individuals are thus expected to be differentially sensitive to organizational stimuli, and if variation in organizational and personal variables is observed simultaneously, we might expect to find interactions. For example, ability to perceive subtle variations in standard operating procedures, sources of influence, and status may be more relevant to success as an administrator when information about these organizational characteristics is made available informally, inaccurately, or not at all, than when it is disseminated systematically in bulletins or memos. One source of conflict in organizations stems from uncertainty about organizational goals (March & Simon, 1959). An in-

formation transition system that produces such uncertainty will increase the relevance of the personality characteristic that has been called uncertainty absorption—the capacity to set premise for the decision of subordinates and peers despite the ambiguity of the environment (Guetzkow & Forehand, 1961). There could be many more examples; these serve to illustrate the interaction hypotheses that might be investigated by varying stimulus characteristics of organizational climate.

Constraints upon Freedom

The stimulus conditions of the organizational environment place bounds upon the set of behaviors that might be selected. Within these bounds, there are additional restrictions imposed by the formal and informal characteristics of an organization, in the form of routine, institutionalized procedures, or of intended or unintended evaluation criteria that place a premium on particular kinds of solution to problems. Hauge (1964), for example, argues that specialization is a threat to diversity and freedom because it "spells a narrowing of the alternatives."

The constraining effects of environment have been emphasized by theorists concerned with creativity. Rogers (1954), for example, maintained that psychological freedom, psychological safety, and a nonevaluative atmosphere are essential conditions in the environment for creativity. Studies of organizational conditions under which scientists work indicate that the extent of constraint varies among organizations, and that it has an effect upon the satisfactions and productivity of the scientists (Vollmer, 1962, 1963).

The origins of such constraints often lie in rules and procedures designed to enable the efficient conduct of complex organizational activities (Cyert & March, 1963) and for keeping delegated decision on the track of organizational goals (Strauss, 1963). It is clear that such procedures are necessary, but also that they can have unintended consequences of ruling out certain kinds of behavior (March & Simon, 1959). There is evidence of variations in the degree to which organizations rely upon rules and procedures (Dill, 1958; Forehand & Guetzkow, 1962), and of concomitant variation in degree of autonomy of individuals (Dill, 1958) and of satisfactions (Herzberg, Mausner, & Snyderman, 1959), and for the general proposition that "programmed activity drives out unprogrammed activity" (March & Simon, 1959).

The most notable effect of such constraints, however, may be interactive in nature. There is evidence from differential studies (e.g., Crites, 1961) that individuals vary in their preference for constraining versus free environments. Thus the performance and satisfaction of individuals would depend upon the particular combination of person and environment. A number of studies provide evidence of interactions between environmental freedom and personal traits. Forehand and Guetzkow (1962) found different patterns of correlations between test-measured characteristics and ratings of innovative behaviors in different climates. With supervisors' ratings as a criterion interpersonal variables such as dominance and need for aggression were most highly related in rule-centered organizations, while cognitive variables such as flexibility and sensitivity to problems were more highly related in group-centered organizations. Barnes (1960) found differential relations between satisfaction and productivity in a closed versus an open system of work-group interrelations. Lodahl and Porter (1961) found that in small

(4–13 men) industrial work groups, homogeneity of members with respect to supervisory attitudes was related to productivity when necessity for work-group cooperation was high, but not when it was low or moderate. Thus, the importance of being like other people in the group depends upon an organizational characteristic.

Another source of constraining influence stems from outside the boundaries of the organization—from the cultural setting in which the organization functions. Haire, Ghiselli, and Porter (1963) found that managers' attitudes toward management in different countries followed a pattern more closely associated with cultural variation than with degree of industrialization, which has frequently been posited as a determinant of attitudes. Thus, managers from Spain and Italy tended to agree more in attitudes toward management—despite their dissimilarity in degree of industrialization—than did, for example, managers from Italy and Denmark whose degrees of industrialization are more similar. Dill (1958) has noted variations in the environment of organizations on a much smaller geographical scale, that is, within countries or even within cities.

The effects of restraint of freedom are not always negative. Blau and Scott (1962) conclude from a review of communication studies, that a hierarchical organization may be more effective than a nonhierarchical organization in facilitating the coordination of work because of the restriction in the flow of communication, while a nonhierarchical organization is more effective in producing new ideas. The establishment of functional roles in a group can also aid the attainment of group goals, while at the same time restricting the freedom of choice of individuals (Golembiewski, 1962b).

Reward and Punishment

Two kinds of effects of stimulus restriction and behavioral constraint have been assumed in the literature. In the first place, they have the result of selecting behaviors simply by permitting some possible responses and precluding others. It has also been assumed that these organizational properties have motivating effects. The human relations emphasis, for example, has assumed that freedom from constraint will generally result in greater satisfaction. Several bases for motivating effects are plausible. First, different organizational properties carry differential opportunity for satisfying the values that employees bring with them into the organization (Vroom, 1960). Secondly, there is growing support for the hypothesis that diversity of stimulation is intrinsically motivating (Fiske & Maddi, 1961). Finally, the mechanism that has received most attention in the literature is the capacity of groups and organizations to reward and punish behavior systematically.

The power of the work group to reward and punish is illustrated in many studies. Among the properties of work groups that may serve a motivating function are cohesiveness (Likert, 1961) and congruence of attitudes (Blau & Scott, 1962). Blau and Scott find evidence that behavior shifts from that which accords with the individual's own values to that which accords with the values of his work group, and interpret such shifts in terms of rewards and punishments provided by the work group.

Properties of larger organizational units also have motivational effects. The findings of Herzberg et al. (1959) suggest that such properties are more likely to have a punishing than a rewarding effect. The factors most often mentioned by the respondents as deter-

minants of dissatisfaction centered about the context of the job—for example, supervision and working conditions—while the factors determining satisfaction were more often characteristics of the job itself—for example, the nature of the work, accomplishment, and responsibility. It should be pointed out that defense mechanisms affecting the perceptions of respondents might account for this finding (Vroom & Maier, 1961).

Among the hypothetically influential properties of organizational reward systems identified by March and Simon (1959, pp. 61 ff.) are: amount of reward, dependence of promotion and monetary reward on performance, the perceived operationality of criteria, and independence of individual rewards. These variables may have both direct and interactive effects. For example, March and Simon proposed (p. 70) that competition among group members will be inversely related to independence of individual rewards; observations of behavior in zero-sum games support this hypothesis (Blake & Mouton, 1961, 1962). Conditions that lead to intragroup competition would be expected to differentially affect persons depending on their own typical level of competitiveness. Blau (1954), in fact, found that personal competitiveness was related to productivity in a competitive work group, but not in a noncompetitive work group.

Dimensions of Organizational Variation

What we have called organizational climate has often been discussed in terms of analogy with individual personality. The validity and implications of the analogy are discussed below, but the analogy is particularly useful in discussing the problem of dimensions. Organizational behavior, like human behavior, is characterized by an overwhelming number of variations. Talking about them requires some way to select and classify them; hence, the search for dimensions—or traits—of organizations. The question of dimensions might be approached empirically in two different ways. First, dimensions might be defined in terms of covariation among many indices of organizational behavior, directly in the tradition of statistical definition of traits. Secondly, organizational variables might be organized according to the effects that they exert, or might be expected to exert, on individuals. These approaches are neither mutually dependent nor mutually exclusive; it is probable that both kinds of analyses will need to be undertaken. Several writers have proposed taxonomies of environmental situations, on the basis of both empirical (Hemphill, 1956; Palmer, 1961) and logical (Golembiewski, 1962b; Sells, 1963b) grounds, while others have reviewed the evidence of dependable effects of specific organizational properties (Thomas & Fink, 1963).

The variables discussed below have been established neither as factors accounting for covariation nor as dependable determinants of behavior; they may be considered as likely candidates for such status in future research. The five variables discussed here—size, organization structure, systems complexity, leadership pattern, and goal directions—were culled from a list of about 30 properties mentioned in studies and discussions of organizational variation.

Size

Thomas and Fink (1963) reviewed 31 studies of effects of group size—with size varying from 2 to 20 members—on group performance, distribution of participation, nature of interaction, group organization, individual performance, conformity, and consensus and member

satisfaction. They point out that size may be considered "phenotypic and really but a correlate of the social and psychological condition capable of producing changes in member and group behavior," and suggest, as intervening variables, resource input (e.g., skills, knowledge), demand input (e.g., members' need for recognition, social interaction), and potential relational complexity.

As one moves from small groups to large organizations, size takes on a different significance. As the number of possible person-to-person interactions increase, the capacity of individuals to form relationships becomes exhausted. A result is the development of small face-to-face subgroups within the large organization, and of pressures toward cohesiveness and adoption of subgroup norms, sometimes at the expense of larger organizational goals (Blau & Scott, 1962; Golembiewski, 1962b; Thomas & Fink, 1963).

The effects of group size upon an individual depend in large measure on the individual's position within the organization. The Thomas-Fink review, as well as studies in applied settings (e.g., Baumgartel & Sobol, 1959; Talacchi, 1960; Thomas, 1959), suggests that satisfaction of work-group members decreases as size of work group increases. Frequently mentioned as factors in this relationship are varying opportunities for participation and for satisfaction of achievement and affiliation needs. In the large organization, the existence of subgroups has the effect of removing the individual from goals of the organization. If the individual is to be evaluated according to production criteria, his evaluated performance may be influenced by conflicts between organizational and subgroup goals (e.g., with regard to quotas). The accoutrements of procedures for efficiency—systems of authority, status, technology, and financial control—place even greater constraints on the individual's opportunity to satisfy personal needs.

For the manager, increasing organizational size may require different sets of skills (Ghiselli, 1963; Haire, 1959). The face-to-face techniques of management must give way to dealing with subgroups and coordinating their outputs. Thus, size may well be a factor influencing correlates of managerial success (Guetzkow & Forehand, 1961).

Structure

Closely related to variables in size, and often varying with them, are variations in the structure of authority and relationships among persons and groups. Evan (1963) summarized evidence that organizations differ widely in the degree of hierarchical organizations of skills, rewards, and authority. There have been many assumptions about behavioral consequences of different kinds of structure, but relatively few empirical attempts to verify them.

Porter and Lawler (1963) examined the proposition most clearly put forward by Worthy (1950) that "flat" (i.e., decentralized) organizations are superior to "tall" (centralized) organizations with respect to satisfaction of employees' needs. They studied responses of over 1,900 managers, classified as belonging to tall, intermediate, and flat organizations, according to the ratio of levels of supervision relative to size. They found no evidence of superiority of flat organizations across all the organizations. There was evidence, however, of interaction between size and shape of the organization: in relatively small organizations (under 5,000 employees), the extent to which managers report their needs to be satisfied was higher for flat than for tall organizations, but in companies with more than

5,000 employees, reported need satisfaction was greater for tall organizations.

Experimental studies, especially those involving communication networks, indicate that satisfaction with job and results are greater in structures with a wider spread of participation. In more centralized structures, satisfaction is highest in the more central positions (Glanzer & Glaser, 1961; Golembiewski, 1962b). There is also evidence that centralized structures are more effective than noncentralized structures for coordinating results, the responsibility for which falls mainly to persons in central positions. These results help to explain why the managers studied by Porter and Lawler (1963) found the tall organization more satisfying in large organizations. Such a structure may give the manager more control over organizational activity and more opportunity for creative contribution than would a flat structure. Porter and Lawler suggest that the effect of structure on need satisfaction might depend upon the individuals position in the organization and that nonmanagers might respond differently.

Another structural variation has been noted by some observers, but has received little research attention; that is, the tendency of the model organizational pattern to shift from a "pyramidal" structure—with workers in the majority at the base—to a "hexagonal structure." The hexagonal structure has blue-collar workers at the base, about equal in number to the top management team, with the largest group, represented in the middle, being professional staff people and those who support them. An implication of the structure of difference of the performance of managers is the problem of the effective supervision of these highly individualistic staff professionals. For the "man in the middle" the difference might be associated with increased conflicts between organizational and professional objectives, decreased opportunity to select assignments and working patterns, and decreased apparent relevance of his work to the functioning of the organization as a whole.

Systems Complexity

The concept of system as a collection of elements functioning together has generated a wide variety of research relevant to organizational behavior. Since analysis of a system implies definition of its elements and quantitative study of their interactions (Peach, 1962), such research can result in unprecedentedly precise descriptions of organizational units (e.g., Chapman et al., 1959; Miller, 1962). While this approach has not yet reached the point of providing comparisons of complex organizational properties and their effects, it appears likely to contribute to that objective.

The systems concept provides several ways of describing environmental variation, and there is little basis for predicting which ones will be useful. One potentially definable variable will serve as an example. Organizations may vary in the complexity of the systems that they employ. Complexity can be defined in terms of the number of components and the number and nature of the interactions among them. It seems likely that variation in systems complexity will be associated with variation in both stimulus availability and behavioral constraint. The part that social elements play in the system offers other opportunities for defining variations in climate (Michael, 1959).

Leadership Style

The various approaches to defining leadership traits (Petrullo & Bass, 1961) also can be applied to the description of

organizations, since significant organizational properties are controlled by persons in leadership positions. Thus a tentative measure of aspects of climate may be obtained by simply taking personality measurements of leaders. For example, Vroom and Mann (1960) and Haythorn, Couch, Haefner, Langham, and Carter (1956) studied follower behavior as a function of leaders' degree of authoritarianism.

The studies of participative management practices, summarized in preceding pages, also illustrate the feasibility of defining leadership style as a dimension of organizational climate. These studies, for the most part, created the contrast experimentally by training or instructing leaders. A further question would ask whether the influence of such variation extends past the individual—whether the variation is truly characteristic of organizations rather than simply of certain individuals. There is some evidence that organizations can be reliably described in terms of typical leadership practices (Baumgartel, 1957; Forehand & Guetzkow, 1962) and that such variation has systematic effects. There is, however, insufficient data as yet to make a conclusion about the generality of such findings.

Goal Directions

Variation in organizational goals is obvious to observers of organizations. Such variation has provided one basis for classifying organizations, for example, as business, government, or philanthropic organizations. Even among business organizations, of which the profit motive might be the most characteristic feature, there is variation with regard to the relative weight placed on subsidiary goals (e.g., dominance of market, avoidance of conflict with government). That such variation would affect behavior of organization members should be expected, in view of the role played by organizational goals in defining the aspects of behavior to be rewarded and punished.

The range of variation in factors affecting behavior may not be as great as that of overall organizational objectives. The subgoals chosen to accomplish the overall objectives (e.g., efficiency, cooperation) may be quite similar despite wide differences in the ultimate goals, and hence the pressures on individuals may be similar. Evidence of relationships between organizational goals and participants' behavior is scarce, partially because of the difficulty of specifying such goals unambiguously and partially because the influence of goals cannot be separated from other organizational qualities (e.g., leadership style). Studies of college environments (Thistlethwaite, 1960) indicate that such effects may exist, however. It was possible to identify schools associated with students' motives to seek higher education in humanities and social sciences as contrasted with those associated with motivation to seek advanced degrees in the natural and biological sciences. The former are characterized by faculty affiliation, enthusiasm and emphasis on achievement, humanism and independence; the latter by a lack of emphasis on compliance. Goals of work groups defined on more specific and more structured situations have received more intensive research attention, and the mechanisms by which conformity pressures are exerted have been explored (Blau & Scott, 1962; Lawrence & Smith, 1955).

Organizational goals may also interact with personal characteristics, particularly the motives of individual organization members. Such interaction may be manifested in several ways: (*a*) The extent to which the individual perceives and understands the organiza-

tion's goals may depend upon his own skills and attitudes (Vroom, 1960). (*b*) The individual who, for one reason or another, responds to his own goals, ignoring those of his organization, can succeed to the extent that his goals coincide with those of the organization. (*c*) The individual who responds both to his own and to his organization's goals faces the possibility of conflict, depending upon what his own goals are. The particular form of the conflict and the attempted resolution both depend in large part on personal factors.

Discussion

This paper has focused mainly upon opportunities for the operational study of environmental variation afforded by organizational research. We have on several occasions drawn an analogy between the climate of an organization and the personality of an individual in order to bring attention to issues in the study of environment. There often appear in the literature, particularly the management literature, suggestions and assumptions that the analogy can be taken more literally, that climate may be treated as a construct, and the "personality of an organization" identified and dealt with.

It has sometimes been suggested that the matching of organizational and individual characteristics would maximize both organizational effectiveness and individual satisfaction. Such a conclusion is suggested by theories and evidence of interactive effects. If the suggestion is implemented mechanically, however, it does not allow room for change, either of the organization or the person. The matching strategy may hinder both the organization and the person from adapting readily to new situations—the former by the inbreeding of inflexibility, and the latter by the limitation placed on the individual's range of experience (McMurry, 1958).

A number of social scientists have suggested a different kind of application of the climate concept: the achievement of organizational change by means of changing the climate. Leavitt (1964) reviews a number of plans for organizational change that focus on "people variables" in the organizational system (in contrast with plans focusing on structure, technology, and task variables). The most influential of these Leavitt terms "power equalization" approaches. The power equalization approaches include client-centered therapy, "T-group" training, and economic distribution devices like the Scanlon Plan. They share an emphasis upon pushing the power and responsibility for decision making as far down in the organizational hierarchy as possible, establishing relationships of mutual trust and authenticity, and keeping all channels of communication open. They also share assumptions that employees, in general, will be more satisfied and more productive when working under such conditions; individual differences in response to climate are not taken into account. Finally, the climate concept has been invoked to account for differences in results of management training programs (Fleishman, 1953; Guetzkow, Forehand, & James, 1962).

These and similar suggestions are promising leads for research and strategy. They rest on the assumption that there is such a thing as a general atmosphere, personality, or climate within an organization. Several implications of this assumption will need to be investigated before the usefulness of climate as a construct is established.

1. Identification of comparable organizational units: According to the personality analogy, organizational climate is a concept based upon covariation of

individual differences among organizations. The identification of what organizational units are to serve the role of persons in this model is a complex problem. An individual is a member of many organizations—a face-to-face work group, department, divisions, etc.—which coexist and interact within a larger organizational unit. Environmental characteristics need not covary systematically between levels of a single organization; a department may be rule centered even if the company is not, and vice versa. When a group of organizations are chosen for study some effort to establish their comparability is needed, unless the study deliberately focuses on variation across levels of organization. One possible approach when individual subjects are involved is to focus on an individual's superior and all of the superior's subordinates or the superior's superior and all of his subordinates, etc. In this way, one can obtain comparable organizational units from widely ranging organizations (Forehand, 1963).

2. Homogeneity within the organizational unit: The description of a company personality implies a degree of homogeneity within the company that may be seldom found. The definition of a climate dimension requires evidence that: the objective determinants of the dimension are applicable to all subunits, and the dimension is perceived comparably by individuals in the subunits.

3. Relative permanence: One of the major ambiguities in definitions of climate has resulted from failure to come to grips with the problem of permanence. The term has been used to refer to organizational properties defined by a given leader and thus sensitive to normal personnel change. On the other hand, it sometimes refers to properties that have become traditionalized through policy, procedures, or objective constraints. If climate is to have meaning as a construct, with meaning distinguished from that of leadership, a criterion of relative permanence would seem appropriate to its definition.

4. Mode of combination of organizational properties: As the preceding discussion has stated, present evidence permits no definitive description of dimensions of organizational variation. The attempt to define the climate of an organization raises a different and more complex question about which there is available no evidence: the question of the way in which dimensions are best combined to describe a particular organization. A meaningful combination may be a linear one, but it seems more likely that it will be a pattern or configuration.

The usefulness of the procedures developed for studying climate need not depend upon the establishment of climate as generally applicable theoretical construct, for the procedures are applicable to the much more general problem of specification of the environment. We have mentioned a number of hypotheses and research approaches toward this end in this paper, and they need not be repeated. They may be summarized in terms of two directions for research that seem especially salient to the problem of studying environmental variation.

1. Measurement: As we have seen, there is no lack of ingenious ideas for the measurement of organizational properties. It is seldom clear, however, that two investigators using the same names for their variables are talking about the same, or even correlated, dimensions. It is for this problem that borrowing from the techniques of differental psychology seems most appropriate. Two approaches can be suggested. The attempt to define dimensions of organizational variation via factor analyses

(Hemphill, 1956; Katzell et al., 1961; Palmer, 1961) has yielded results encouraging extensions to a wider range of organizations and organizational properties. A second differential concept that needs emphasis is that of validity, particularly convergent and discriminant validity (Campbell & Fiske, 1959). The multitrait multimethod matrix recommended by Campbell and Fiske for validation of personality tests also provides a model for validating climate measures. Rule centeredness, for example, measured by participants' perceptions, ought to correlate more highly with the same variable measured by an objective count than with a different variable measured by perceptions. Otherwise the perceptual process, rather than the organization's properties, accounts for variance in the measure. Studies of this sort would not only validate the measures for use in future experiments, but would also contribute information about the nature of effective organizational variation.

2. Interaction studies: Many approaches to the study of behavior assume that either the person or the environment varies, but not both at the same time. Both the theoretical disadvantage and the practical inutility of such an approach have been demonstrated (Cronbach, 1957; Helson, 1964). Studies of organizations make possible the variation of both sets of variables, and thus provide not only a greater understanding of the functions of organizations, but also a chance to test psychological hypotheses about the interaction of the individual and his environment.

REFERENCES

ABELSON, R. P. Commentary: Situational variables in personality research. In S. Messick & J. Ross (Eds.), *Measurement in personality and cognition.* New York: Wiley, 1962.

ARGYRIS, C. Some problems in conceptualizing organizational climate: A case study of a bank. *Admin. Sci. Quart.,* 1958, 2, 501–520.

ARGYRIS, C. *Interpersonal competence and organizational effectiveness.* Homewood, Ill.: Dorsey Press, 1962.

ARNOULT, M. D. The specification of a "social" stimulus. In S. B. Sells (Ed.), *Stimulus determinants of behavior.* New York: Ronald Press, 1963.

BARNES, L. B. *Organizational systems and engineering groups.* Boston: Harvard University, Graduate School of Business Administration, 1960.

BARTON, A. H. *Organizational measurement and its bearing on college environments.* New York: College Entrance Examination Board, 1961.

BASS, B. M. Experimenting with simulated manufacturing organizations. In S. B. Sells (Ed.), *Stimulus determinants of behavior.* New York: Ronald Press, 1963.

BAUMGARTEL, H. Leadership style as a variable in research administration. *Admin. Sci. Quart.,* 1957, 2, 344–360.

BAUMGARTEL, H., & SOBOL, R. Background and organization factors in absenteeism. *Personnel Psychol.,* 1959, 12, 431–443.

BLAKE, R. R., & MOUTON, JANE S. Loyalty of representatives to in-group positions during interpersonal competition. *Sociometry,* 1961, 24, 177–183.

BLAKE, R. R., & MOUTON, JANE S. Comprehension of points of communality in competing solutions. *Sociometry,* 1962, **25**, 56–63.

BLAU, P. M. Co-operation and competition in a bureaucracy. *Amer. J. Sociol.,* 1954, **59**, 530–535.

BLAU, P. M., & SCOTT, W. R. *Formal organizations.* San Francisco, Calif.: Chandler 1962.

BRUNER, J. Social psychology and perception In Eleanor Maccoby, T. Newcomb, & E Hartley (Eds.), *Readings in social psychology.* (3rd ed.) New York: Holt, Rinehart, & Winston, 1958.

BRUNSWIK, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.,* 1959, **56**, 81–105.

CHAPMAN, R. L., KENNEDY, J. L., NEWELL, A., & BIEL, W. C. The Systems Research Laboratory's air defense experiments. *Mgmt. Sci.,* 1959, **5**, 250–269.

COCH, L., & FRENCH, J. R. P. Overcoming resistance to change. *Hum. Relat.*, 1948, **1**, 512–532.

COHEN, K. J., & RHENMAN, E. The role of management games in education and research. *Mgmt. Sci.*, 1961, **7**, 131–166.

COSTELLO, T. W., & ZALKIND, S. S. *Psychology in administration: A research orientation.* Englewood Cliffs, N. J.: Prentice-Hall, 1963.

CRITES, J. O. Factor analytic definitions of vocational adjustment. *J. appl. Psychol.*, 1961, **45**, 330–337.

CRONBACH, L. J. The two disciplines of scientific psychology. *Amer. Psychologist*, 1957, **12**, 671–684.

CYERT, R. M., & MARCH, J. G. *A behavioral theory of the firm.* Englewood Cliffs, N. J.: Prentice-Hall, 1963.

DAY, R. C., & HAMBLIN, R. L. Some effects of close and punitive styles of supervision. Technical Report No. 8, 1961, Contract Nonr 816(11), Office of Naval Research.

DEARBORN, D. C., & SIMON, H. A. Selective perception: A note on the departmental identification of executives. *Sociometry*, 1958, **21**, 140–144.

DILL, W. R. Environment as an influence on managerial autonomy. *Admin. Sci. Quart.*, 1958, **2**, 409–443.

DILL, W. R., HILTON, T. L., & REITMAN, W. R. *The new managers.* Englewood Cliffs, N. J.: Prentice-Hall, 1962.

EVAN, W. M. Indices of the hierarchical structure of industrial organizations. *Mgmt. Sci.*, 1963, **9**, 468–477.

FISKE, D. W., & MADDI, S. (Eds.) *Functions of varied experience.* Homewood, Ill.: Dorsey Press, 1961.

FLEISHMAN, E. A. Leadership climate, human relations training and supervision. *Personnel Psychol.*, 1953, **6**, 205–222.

FLEISHMAN, E. A., & HARRIS, E. F. Patterns of leadership behavior related to employee grievances and turnover. *Personnel Psychol.*, 1962, **15**, 43–56.

FOREHAND, G. A. Assessments of innovative behavior: Partial criteria for the assessment of executive performance. *J. appl. Psychol.*, 1963, **47**, 206–213.

FOREHAND, G. A., & GUETZKOW, H. Characteristics related to innovative administrative behavior in varying organizational climates. In, *Education for innovative behavior in executives.* (Final report, Cooperative Research Project No. 975) Washington, D. C.: United States Office of Education, 1962.

GELLERMAN, S. W. *People, problem and profits.* New York: McGraw-Hill, 1960.

GHISELLI, E. E. Managerial talent. *Amer. Psychologist*, 1963, **18**, 631–642.

GIBSON, J. J. The concept of the stimulus in psychology. *Amer. Psychologist*, 1960, **15**, 694–703.

GILMER, B. v. H. *Industrial psychology.* New York: McGraw-Hill, 1961.

GLANZER, H., & GLASER, R. Techniques for the study of group structure and behavior: II. Empirical studies of the effects of structure in small groups. *Psychol. Bull.*, 1961, **58**, 1–27.

GOFFMAN, E. *The presentation of self in everyday life.* New York: Anchor Books, 1959.

GOLEMBIEWSKI, R. T. *Behavior and organization: O & M and the small group.* Chicago: Rand-McNally, 1962. (a)

GOLEMBIEWSKI, R. T. *The small group: An analysis of research concepts and operations.* Chicago: Univer. Chicago Press, 1962. (b)

GREGG, L. W. Engineering psychology. In B. v. H. Gilmer, *Industrial psychology.* New York: McGraw-Hill, 1961.

GUETZKOW, H. (Ed.) *Simulation in social science: Readings.* Englewood Cliffs, N. J.: Prentice-Hall, 1962.

GUETZKOW, H., & BOWES, A. E. The development of organizations in the laboratory. *Mgmt. Sci.*, 1957, **3**, 380–402.

GUETZKOW, H., & FOREHAND, G. A. A research strategy for partial knowledge useful in the selection of executives. In R. Tagiuri (Ed.), *Research needs in executive selection.* Boston: Harvard University, Graduate School of Business Administration, 1961.

GUETZKOW, H., FOREHAND, G. A., & JAMES, B. J. An evaluation of educational influence on executive judgment. *Admin. Sci. Quart.*, 1962, **6**, 483–500.

HAIRE, M. Biological models and empirical histories of the growth of organizations. In M. Haire (Ed.), *Modern organization theory.* New York: Wiley, 1959.

HAIRE, M., GHISELLI, E. E., & PORTER, L. W. Cultural patterns in the role of the manager. *Industr. Relat.*, 1963, **2**, 95–117.

HAUGE, G. *The individual in the modern economic world.* Pittsburgh, Pa.: Carnegie Press, 1964.

HAYTHORN, W., COUCH, A., HAEFNER, D., LANGHAM, P., & CARTER, L. F. The behavior of authoritarian and equalitarian personalities in groups. *Hum. Relat.*, 1956, **9**, 58–74.

HELSON, H. Current trends and issues in adaptation-level theory. *Amer. Psychologist*, 1964, **19**, 26–38.

HELSON, H., BLAKE, R. R., & MOUTON, JANE S. Petition-signing as adjustment to situa-

tional and personal factors. *J. soc. Psychol.*, 1958, **48**, 3–10.

HEMPHILL, J. K. *Group dimensions: A manual for their measurement.* Columbus, O.: Ohio State University, Bureau of Business Research, 1956.

HEMPHILL, J. K., GRIFFITHS, D. E., & FREDERIKSEN, N. *Administrative behavior: A study of the principal in a simulated elementary school.* New York: Columbia University, Teachers College, 1962.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, BARBARA B. *The motivation to work.* New York: Wiley, 1959.

KATZELL, R. A., BARRETT, R. S., & PARKER, T. C. Job satisfaction, job performance, and situational characteristics. *J. appl. Psychol.*, 1961, **45**, 65–72.

KIDD, J. S., & CHRISTY, T. Supervisory procedures and work-team productivity. *J. appl. Psychol.*, 1961, **45**, 388–392.

LAWRENCE, L. C., & SMITH, P. C. Group decision and employee participation. *J. appl. Psychol.*, 1955, **39**, 334–337.

LEAVITT, H. J. Toward organizational psychology. In B. v. H. Gilmer (Ed.), *Walter Van Dyke Bingham Memorial Program.* Pittsburgh, Pa.: Carnegie Institute of Technology, 1962. (a)

LEAVITT, H. J. Unhuman organization. *Harv. Bus. Rev.*, 1962, **40**, 90–98. (b)

LEAVITT, H. J. Applied organizational change in industry: Structural, technological and humanistic approaches. In J. G. March (Ed.), *Handbook of organizations.* Chicago: Rand-McNally, 1964.

LEAVITT, H. J., & BASS, B. M. Organizational psychology. *Annu. Rev. Psychol.*, 1964, **15**, 371–398.

LEWIN, K., LIPPITT, R., & WHITE, R. K. Patterns of aggressive behavior in experimentally created "social climates." *J. soc. Psychol.*, 1939, **10**, 271–299.

LIKERT, R. *New patterns of management.* New York: McGraw-Hill, 1961.

LODAHL, T. M., & PORTER, L. W. Psychometric score patterns, social characteristics, and productivity of small industrial work groups. *J. appl. Psychol.*, 1961, **45**, 73–79.

MCMURRY, R. N. Recruitment, dependence, and morale in the banking industry. *Admin. Sci. Quart.*, 1958, **3**, 87–117.

MARCH, J. G., & SIMON, H. A. *Organizations.* New York: Wiley, 1959.

MICHAEL, D. N. The social environment. *Operat. Res.*, 1959, **7**, 506–523.

MILLER, R. B. Task description and analysis. In R. M. Gagné (Ed.), *Psychological principles in systems development.* New York: Holt, Rinehart, & Winston, 1962.

MORSE, NANCY, & REIMER, E. The experimental change of a major organization variable. *J. abnorm. soc. Psychol.*, 1956, **52**, 120–129.

MURRAY, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.

NEIMARK, EDITH D., & ROSENBERG, S. The effect of "social" discriminative cues. *J. exp. Psychol.*, 1959, **58**, 302–311.

PACE, C. R. Implications of differences in campus atmosphere for evaluation and planning of college programs. In R. L. Sutherland, W. H. Holtzman, E. A. Koile, & B. K. Smith (Eds.), *Personality factors on the college campus.* Austin, Tex.: Hogg Foundation for Mental Health, 1962.

PACE, C. R., & STERN, G. C. An approach to the measurement of psychological characteristics of college environments. *J. educ. Psychol.*, 1958, **49**, 269–277.

PALMER, G. J. Test of a theory of leadership and organization behavior with management gaming. Second annual report, 1961, Louisiana State University, Contract Nonr 1575 (05), Office of Naval Research, Group Psychology Branch.

PEACH, P. What is system analysis? In H. W. Karn & B. v. H. Gilmer (Eds.), *Readings in industrial and business psychology.* New York: McGraw-Hill, 1962.

PELZ, D. C. Leadership within a hierarchical organization. *J. soc. Issues*, 1951, **7**, 47–63.

PETRULLO, L., & BASS, B. M. (Eds.) *Leadership and interpersonal behavior.* New York: Holt, Rinehart, & Winston, 1961.

PORTER, L. W., & LAWLER, E. E. The effects of "tall" vs. "flat" organizations on managerial satisfactions. Paper read at American Psychological Association, Philadelphia, September 1963.

ROGERS, C. R. Toward a theory of creativity. *Etc.*, 1954, **4**, 249–260.

ROSENBAUM, M. The effect of stimulus and background factors in the volunteering response. *J. abnorm. soc. Psychol.*, 1956, **53**, 118–121.

ROSENBERG, S., & HALL, R. L. The effects of different social feedback conditions upon performance in dyadic teams. *J. abnorm. soc. Psychol.*, 1958, **58**, 271–277.

SELLS, S. B. Dimensions of stimulus situations which account for behavior variance. In S. B. Sells (Ed.), *Stimulus determinants of behavior.* New York: Ronald Press, 1963. (a)

SELLS, S. B. An interactionist looks at the environment. *Amer. Psychologist*, 1963, **18**, 696–702. (b)

SELLS, S. B. (Ed.) *Stimulus determinants of behavior.* New York: Ronald Press, 1963. (c)

SHERMAN, P. A., KERR, W., & KOSINAR, W. A study of accidents in 147 factories. *Personnel Psychol.*, 1957, **10**, 43–51.

STANTON, E. S. Company policies and supervisors' attitude toward supervision. *J. appl. Psychol.*, 1960, **44**, 22–26.

STRAUSS, G. Some notes on power equalization. In H. J. Leavitt (Ed.), *The social science of organization: Four perspectives.* Englewood Cliffs, N. J.: Prentice-Hall, 1963.

STROTHER, G. B. Problems in the development of a social science of organizations. In H. J. Leavitt (Ed.), *The social science of organizations: Four perspectives.* Englewood Cliffs, N. J.: Prentice-Hall, 1963.

TAFT, R. The ability to judge people. *Psychol. Bull.*, 1955, **52**, 1–21.

TAGIURI, R. (Ed.) *Research needs in executive selection.* Boston: Harvard University, Graduate School of Business Administration, 1961.

TAGIURI, R., & PETRULLO, L. (Eds.) *Person perception and interpersonal behavior.* Stanford: Stanford Univer. Press, 1958.

TALACCHI, S. Organization size, individual attitudes, and behavior: An empirical study. *Admin. Sci. Quart.*, 1960, **5**, 398–420.

THISTLETHWAITE, D. L. College press and student achievement. *J. educ. Psychol.*, 1959, **50**, 183–191.

THISTLETHWAITE, D. L. College press and changes in study plans of talented students. *J. educ. Psychol.*, 1960, **51**, 222–234.

THOMAS, E. J. Role conceptions and organizational size. *Amer. sociol. Rev.*, 1959, **24**, 30–37.

THOMAS, E. J., & FINK, C. F. Effects of group size. *Psychol. Bull.*, 1963, **60**, 371–384.

VOLLMER, H. M. A preliminary investigation and analysis of the role of scientists in research organizations. I. *USAF OSR tech. Rep.*, 1962, Task No. 37707, Project No. 9778.

VOLLMER, H. M. Adaptations of scientists in an independent research organization: A case study. IIa. *USAF ORD tech. Rep.*, 1963, Task No. 37707, Project No. 9778.

VROOM, V. H. The effects of attitudes on perception of organizational goals. *Hum. Relat.*, 1960, **13**, 229–240.

VROOM, V. H., & MAIER, N. R. F. Industrial social psychology. *Annu. Rev. Psychol.*, 1961, **12**, 413–446.

VROOM, V. H., & MANN, F. C. Leader authoritarianism and employee attitudes. *Personnel Psychol.*, 1960, **13**, 125–140.

WORTHY, J. C. Factors influencing employee morale. *Amer. sociol. Rev.*, 1950, **15**, 169–179.

ZANDER, A., & CURTIS, T. Effects of social power on aspiration level and striving. *J. abnorm. soc. Psychol.*, 1962, **64**, 63–74.

(Received March 16, 1964)

Psychological Bulletin
1964, Vol. 62, No. 6, 383–393

CONDITIONING OF VERBALIZATION: A REVIEW[1]

JUANITA H. WILLIAMS

Veterans Administration Center, Bay Pines, Florida

Recent research in the conditioning of verbal behavior offers support to a learning-theory interpretation of changes that may occur in psychotherapy. Manipulation of awareness states, use of verbal conditioning as an independent variable, and effect of the reinforcement history of S in verbal conditioning experiments are being explored in relevance to controlled and measurable modification of verbal behavior. Evidence is accumulating that the generalization of an experimentally acquired verbal response is a function of the common cue-producing properties of the several classes of which it may be a member.

Psychotherapeutic practice is presently undergoing a rather remarkable reevaluation owing to results being reported by researchers utilizing techniques based on modern learning theory. In practice and research, methods based on classical and operant conditioning theory have been devised to modify many kinds of unadaptive behavior, and a growing body of literature attests to their efficacy and to their value as objective, replicable, and measurable agents of change.

Another aspect of theoretical interest, however, concerns the modification of behavior which can occur in situations other than those so structured and controlled. In traditional psychotherapy, for example, the therapist deals, for the most part, with verbal responses emitted by the client. Since in a two-person situation such as this each individual acts as a set of stimuli which elicit responses from the other, responses which are positively or negatively reinforced, even without awareness, it is clear that this can be construed as a learning situation.

From this point of view, the role of the therapist as a source of reinforcement contingent upon classes of client verbal behavior has come under scrutiny and many studies have appeared which illuminate this situation and which support a learning theory approach to psychotherapy.

This review is an attempt to survey the recent research in the area of the modification of verbal behavior within an operant conditioning paradigm, where the reinforcer is some form of social approval mediated by another person.

At the outset, however, it is necessary to deal with an important theoretical issue which has been raised by Dulaney (1961). This issue is whether it is justifiable to call the behavior changes reported in studies of verbal conditioning operant conditioning, in the sense of being a simple analog of animal conditioning to bar pressing or disc pecking. In his experiment, which was a partial replication of Greenspoon's (1955), subjects were verbally reinforced for plural nouns in the last four of five blocks of freely emitted words. Analysis of the data showed that, while 35 of

[1] This review represents a modification of a portion of a doctoral dissertation which was presented to the faculty of Temple University in partial fulfillment of the requirements for the degree of doctor of philosophy.

the 43 experimental subjects increased their mean frequency of plural noun response in the four reinforced blocks compared to the nonreinforced block, there was no significant difference in trend over the five blocks. Analysis of answers to questions asked subjects after each block revealed that no subject could state that the contingency for reinforcement was a response of a plural noun. However, 34 reported that the experimenter was studying their associations, and of these, 11 reported that "umhmm" was a reinforcement for staying in the same category. When experimental subjects were categorized into groups labeled reinforcement for association, associative hypothesis alone, and no associative hypothesis, it was found that the first groups showed a highly significant learning effect, the second a modest one, and the third were no different from controls. This raises the question of what it is that is learned in such an experiment. Although the probability of the critical response was increased, subjects did not learn to produce plural nouns but rather formed a hypothesis that continued associating in a series, for example, "jewels" or "vegetables" would be rewarded, and thus developed a set to stay in such a category. Dulaney proposes that the effect could then be ascribed to the mediation of hypotheses that cue prior verbal habits. In Salzinger's (1959) list of the controls for verbal operant conditioning, verbal hypotheses are not included, which raises the question, can the learning effect of subjects who verbalize an associative hypothesis be described as conditioning? Dulaney feels that verbal operant conditioning may eventually turn out to be human problem solving, mediated by hypotheses and transfer, but that the discovery of mediating mechanisms does not alter the fit of empirical operant principles, since a free operant came under the control of a contingent stimulus.

Greenspoon (1962) has raised two other related issues having to do with the legitimacy of subsuming all procedures used in this type of study under the rubric of verbal conditioning. One of these is that it is clear that in many of the experiments reported to date, the subject does not acquire any new verbal responses; nor does he freely emit critical responses, since they are often presented to him in a rather highly structured situation. The other issue concerns the importance of the generic nature of the response; as Skinner (1935) has said, it is necessary to conceive of classes of responses, whose members have common characteristics. Some experiments, for example, have presented a situation where a specific word, such as a personal pronoun, was reinforced; while these studies have demonstrated the acquisition of the response, they do not fit within the operant conditioning paradigm owing to the uniqueness of the single response, which does not provide opportunity for generalization. It is Greenspoon's conclusion, however, that if the paradigm is characterized by a shift in probability of a response, then verbal conditioning research may be considered within the paradigm. Thus, while these considerations raise interesting questions as to the justification for considering verbal operant conditioning experiments analogous to those operant conditioning studies on the acquisition of responses by infrahumans, it is assumed here that verbal behavior is acquired by and is modifiable by the same laws of learning that control other behavior; in particular, that verbal response probabilities can be changed by the systematic introduction of stimuli which are contingent upon the selected response or class of responses.

Althought verbal learning and verbal behavior have been the subject of experimental effort at least since 1885 when Ebbinghaus published *On Memory,* it was Thorndike (1935) who first formulated the hypothesis that verbal behavior could be modified by the administration of verbal reward and punishment. His attention was chiefly given, however, to showing the efficacy of reward over punishment in confirming responses by strengthening connections rather than to the control of verbal behavior per se.

The conditioning, extinction, and generalization of verbal behavior has received attention for several decades (Humphreys, 1939; Johnson, 1944; Razran, 1949); but the current interest in the experimental modification of verbal behavior following a Skinnerian paradigm appears to have originated with Greenspoon (1951), who was able to demonstrate that four different reinforcements, verbal approval and disapproval, a light and a tone, changed the probability of a response class of plural nouns. In recent years, numerous studies (e.g., Gross, 1959; Reidy, 1958; Taffel, 1955; Verplanck, 1955) have confirmed the finding that various verbal response classes can be caused to increase in frequency by the introduction of some type of reinforcing contingency, usually verbal approval; however, lights (Sidowski, 1954), buzzer (Greenspoon, 1951), and a bell tone (McNair, 1957) have been reported to give positive results, as well as such nonverbal social reinforcers as head-nodding, smiling, or postural shifts (Wickes, 1956).

While the majority of studies report successful conditioning of many types of response classes, others, using similar classes and reinforcing stimuli, report negative results. Spielberger and DeNike (1962) found that unaware subjects did not differ from controls in rate of emission of plural nouns, and suggest that positive results in this type of study may be artifacts of the failure to detect awareness in some experimental subjects. Mandler and Kaplan (1956), also replicating the Greenspoon study with negative results, questioned their subjects regarding their interpretation of the reinforcer "um-hmm." Those who identified it as a positive reinforcer had shown evidence of conditioning, while those holding a negative interpretation had shown a decrease in emission of the critical response.

The issue of the role of awareness, usually defined as subject's ability to verbalize under questioning of various sorts the relationship between his responses and the reinforcer employed, has been dealt with by many investigators. Adams (1957), in an extensive review of studies on learning and awareness, has concluded that evidence for learning without awareness is equivocal. He points out that the criteria for such learning in studies of verbal behavior are different in different investigations, and that there is little uniformity as to what constitutes awareness or how to measure it. He suggests that the number of subjects judged aware is in part a function of the number and type of questions asked them after completion of the conditioning procedure. The usual method for judging awareness in these earlier studies required subjects to complete open-ended questionnaires after the conditioning procedure (Greenspoon, 1955; Mandler & Kaplan, 1956; Taffel, 1955). Krasner (1958), in a review of 31 studies of the conditioning of verbal behavior, has reported that approximately 5% of all subjects in the combined studies were aware in the judgment of the various experimenters. However, the adequacy of the methodology and the

validity of the criteria for awareness in these investigations are questionable; the evidence (Dulaney, 1961) that acquisition of a verbal response occurs through the mediation of verbal hypotheses implies that awareness on some level is always a factor where the conditioning effect is observed.

Postman and Sassenrath (1961), in a survey of studies of incidental learning and of learning without awareness, take issue with Adams' viewpoint. They conclude that learning without awareness does occur, and that verbal rewards and punishments have a significant effect under conditions of incidental learning. It is unnecessary to assume that behavior modification must be preceded by a correct understanding of the environmental contingencies; therefore, the question of whether the action of verbal rewards and punishments is always or never automatic is not a profitable one.

In a study of the effects of awareness on conditioning, Levin (1961) found that lengthening the awareness interview increased the number of subjects judged aware, and that the aware subjects accounted for the differences between experimental and control group performance. Krasner, Weiss, and Ullmann (1961), using subjects who had been exposed to academic material on verbal conditioning and who therefore were not naive, studied awareness first as a dependent variable in a procedure typical of studies reported earlier. They then treated it as an independent variable, manipulating it by using different sets of instructions in a preconditioning interview. They conclude that the specific effect that awareness will have on conditioning depends on subject-determined variables, such as the subject's emotional attitude toward the experimenter, and feel that, taken alone, awareness is a concept of dubious validity in verbal conditioning studies. Southwell (1962) has criticized this study on the basis of the definition of awareness employed. Data from other studies (Ekman, Krasner, & Ullmann, 1963; Kanfer & Marston, 1962; Simkins, 1963; Spielberger, 1962; Spielberger, Levin, & Shepherd, 1962) support the conclusions that awareness is a function of preconditioning instructions, discriminability of critical response and reinforcement, personality interaction, and atmosphere, and that these variables can be controlled to influence reported awareness. It is clear that this issue is coming to assume an importance for some investigators which is comparable to that of conditioning itself; and though a strictly Skinnerian view would not admit that such private events should have systematic status among variables controlling behavior (Skinner, 1953), another view (Krasner, 1962), more clinically oriented, holds that awareness as a manipulable variable can be used in the investigation of such social influence situations as placebos, attitude control, and psychotherapy. In the latter, it can be viewed as a type of insight, controlled by the therapist and studied as a function of his value system and of his theoretical orientation.

Reviews of the literature (Greenspoon, 1962; Krasner, 1958, 1962; Salzinger, 1959) have laid increasing stress upon the importance of defining the variables operating in such experimental situations. Equivocal and contradictory results reported in these reviews appear to be a function of the complexity of the phenomenon and of the difficulty of controlling all important variables. While the early studies were concerned with demonstrating the acquisition of a verbal response as a function of the type of response and of the type of reinforcement, more

recent workers have investigated such questions as the effect of resistance to extinction as a function of number of acquisition trials (Senko, Champ, & Capaldi, 1961) and of experimenter-subject interaction (Binder, McConnell, & Sjohelm, 1957; Gross, 1959; Krasner, Ullmann, Weiss, & Collins, 1961; Sapolski, 1960). In addition, there is increasing interest in the reinforcement history of the individual, which manifests itself in the state of the subject when he enters the experimental situation (Salzinger, 1959). Attempts to measure the effects of such states on conditionability fall into two general categories. One line of investigation uses various devices to measure, in a sense, the reinforcement history of the subject, defined operationally by his scores on tests of personality variables. The other is concerned with manipulating some intraindividual state experimentally and then observing its effect on conditioning.

Weiss, Ullmann, and Krasner (1960) studied the relationship between responsivity to verbal conditioning and likelihood of hypnotizability, in an attempt to clarity a problem posed by Das (1958). Das had found a significant correlation between classical eyeblink conditioning and hypnotizability, but had noted that a difficulty in interpreting his results was a possible facilitating effect of one treatment on another in the procedure. To handle this, Weiss et al. used an indirect approach, a measure of likelihood of hypnotizability instead of actual hypnotic induction. A significant correlation between the two variables was interpreted by the authors as supporting Das' findings and broadening the generality of the hypothesis, by using verbal operant rather than classical conditioning.

Reidy (1958), hypothesizing that need for approval would act as a drive influencing the acquisition of a verbal response, found that the personality variable had no effect on learning. Results of a similar study (Crowne & Strickland, 1961), however, indicated that need for approval facilitated the conditioning of verbal behavior.

Babladelis (1961) tested the hypothesis that the relationship between autonomy and verbal conditioning effects is linear and negative and found that acquisition of self-statement responses was correlated −.60 with autonomy. This appears to be related to a secondary finding of the Weiss et al. (1960) study, in which a correlation of −.43 was found between conditionability and achievement via independence.

Level of performance in a verbal conditioning task was found to be a function of anxiety, compliance, and defensiveness in a study by Sarason (1958). Using test measures and therapists' ratings, he found that high scores on test anxiety and lack of protection were associated with poor conditioning. Patients rated by therapists as compliant showed a greater learning effect. Sarason concludes that it is important to control for individual personality differences; that subjects' previously learned attitudes and anticipations affect the extent to which they accept reinforcements or use them as a basis for modifying their own behavior.

Other subject variables such as neuroticism (Everstine & Bendig, 1960) and hostility (Campbell, 1960) were found to have an effect on conditionability; but Hetrick and Haas (1962) found no correlation between the variables of ego strength, depression and psychopathy, and performance in a verbal conditioning task. No support for Eysenck's hypothesis on the differential conditionability of extraverts and introverts appeared in results reported

by Das (1961), Das and Mitra (1962), and McDonnell and Inglis (1962).

Krasner, Ullmann, Weiss, and Collins (1961) have utilized some of their findings and others (Ferguson & Buss, 1960; Weiss et al., 1960) to develop an inventory which is related to resistance to conditionability. Consisting of the Achievement via Independence and Susceptibility to Hypnosis scales from the California Psychological Inventory and the MMPI Manifest Hostility scale, it predicts individual responsivity to conditioning.

Although such personality variables cannot be considered to be uncorrelated, their number and variety and the nature of the relationships found furnish evidence of the complexity of the processes underlying conditionability as manifested by acquisition of responses in verbal situations.

The experimental subject not only brings with him into the conditioning situation a history of reinforcement which affects his performance, but he may be exposed systematically to events within the situation which are designed to produce a condition whose affect on his conditionability can be studied. Gewirtz and Baer (1958) studied the effect of a social reinforcer on learning in children under conditions of social deprivation and satiation. They found that children who had been experimentally deprived of social contact responded at a higher rate to the social reinforcer than did children who had been experimentally satiated with such contact. Though the response learned was not a verbal one, the study is relevant in that it indicates that a verbal reinforcer is responsive to depdrivation and satiation of the social drives of the subject.

The effects of emotional atmosphere and of withholding reinforcement were studied (Weiss & Ullmann, 1960) in an investigation of the appropriateness of verbal conditioning for studying interpersonal responsiveness. Emotional words used in the telling of Thematic Apperception Test-like stories were verbally reinforced by the experimenter. Following conditioning, the experimenter interacted with half the subjects in a hostile manner and with the other half in a friendly manner. Within each of these atmosphere conditions, half the subjects underwent extinction trials. All subjects then received additional reinforcement trials, so that both reconditioning and continued reinforcement were studied. The experimental manipulations were found to have a significant effect on responsiveness. A hostile atmosphere and nonreinforcement both resulted in decreased responsiveness. The induced hostility, however, did not affect reconditioning. The authors discuss verbal conditioning as an interpersonal paradigm, suggesting that future research of this type into the nature of interpersonal relationships will need to take into account three variables assumed to be related to personality correlates. These are "expressiveness," which in this study refers to individual differences in the use of emotional words, and is analogous to activity in subhuman subjects; "responsiveness to reinforcement," which is a factor responsible for a decrease in the relationship between operant level and the behavior under study when reinforcement is introduced; and individual differences in response to "situational factors," that is, conditions other than reinforcement introduced by the experimenter.

Hall (1960) found that subjects conditioned faster under conditions of ego orientation to the task. Using male and female subjects, he varied the instructional set, one group receiving ego-oriented instructions, one task-oriented

instructions, and one neither. "Good" was found to be an effective reinforcer for all groups, but the amount of conditioning was a function of the instructional set.

Weiss (1955) studied the effects of set for speed and awareness on learning. The task was to respond to a stimulus word with the first word coming to mind. The four groups were: informed-timed (told the contingency, that living thing words would be reinforced with "good," and set for speed of response); uninformed-timed; uninformed-not-timed; and control. The greatest increment in the critical response was observed for the uninformed-not-timed, suggesting to the author that utilization of the correct principle is impaired by the set for speed. Kanfer and Marston (1962), on the other hand, found that task-relevant information facilitates learning in verbal conditioning experiments. A possible explanation of the Weiss findings, that the informed group showed less learning than the uninformed group, is that the instruction to respond with the first word coming to mind, and the additional information as to the correct class of response, resulted in conflicting response tendencies compared with the uninformed group which had only one set of instructions. Negative results in a study of set have also been reported by Brown and Webb (1960), who were interested in the effect of different sets on the prediction of uncertain alternative events. The three types of set employed in the study did not change the lawful probability of response as outlined by a statistical association theory of learning.

Experimenter-subject interaction has been the subject of a number of studies. Kanfer and Karas (1959), in a systematic manipulation of prior subject-experimenter interaction, found that either praise or criticism was followed by more learning than was noninteraction. Physical characteristics of the experimenter were used by Binder, McConnell, and Sjohelm (1957), as an independent variable to study acquisition of hostile verbs in a sentence-forming task. One experimenter was a small, attractive female, while the other was a husky ex-Marine captain. A trend analysis showed a steeper acquisition curve for the female experimenter's group.

Krasner, Ullmann, Weiss, and Collins (1961) used two male PhDs and a female AB as experimenters in a status-type study, and found that the male experimenters obtained significantly greater use of the critical response, emotional words, in the reinforced trials than during the operant trials, while the female experimenter obtained results in the predicted direction, but not to a significant extent. Forty-four percent of subjects conditioned by male experimenters, and none of those conditioned by the female experimenter, reported some awareness of the contingency.

Characteristics of response classes selected for reinforcement have varied widely. Studies in this area have been concerned with semantic properties of word classes (Staats, 1961) as well as with the categorization of verbal behavior along such dimensions as feeling states and attitudes (Auld & Murray, 1955; Auld & White, 1956). Greenspoon (1962) has noted that some negative results in conditioning studies probably result when the basis of discrimination of the critical response is not sufficiently defined to enable the subject to make the discrimination.

Although many experimental studies have reinforced response classes which consist of single words, as in the Taffel-type procedures, or of classes such as

mother references (Mock, 1957), recent investigators have been interested in the conditioning of more molar types of verbal behavior, such as expressions of opinion (Verplanck, 1955), group therapy behavior (Ullmann, Krasner, & Ekman, 1961), and attitude change (Scott, 1957). As Krasner (1955, 1962) has stated, these attempts have a direct application to psychotherapy, which is more often concerned with the modification of global behavior. While the idea of the therapist as one who controls or manipulates another individual has been the topic of some controversy (Rogers & Skinner, 1956), investigation of the psychotherapeutic situation is of necessity concerned with the exploration of the lawful relationships accounting for any change which can be measured.

Kanfer and McBrearty (1962) used minimal cues such as physical attitude and nondirective verbal responses to reinforce selected topics in a structured interview, and found that such reinforcement resulted in an increase in time spent on such topics. In addition, this study demonstrated that complex verbal classes can be reliably discriminated and coded for investigative purposes. The authors point out, however, that subjects vary considerably in their sensitivity to aspects of the reinforcement and in their interpretation of instructions in such relatively free situations.

Using a stimulus-response concept of personality traits, Staats, Staats, Heard, and Finley (1962) reasoned that a test-inferred trait is measured by items forming a unified class of responses representing sets of behavior that have been learned under the same conditions of reinforcement. Subjects, who were low scorers on the Sociability scale of the Guilford-Zimmerman test, were verbally reinforced for sociable responses. A trend analysis showed that the experimental group responded at a higher rate than a control group. In a similar study, Nuthmann (1957) found that verbally reinforced subjects increased in frequency of self-acceptance statements on a personality test. She points out, however, that it is questionable whether the subjects were responding to a more positive emotional tone or to acceptance of self, which is probably a more abstract concept.

In a study of the conditioning of affective responses, Haas (1962) reported that both positive and negative sentence endings were increased by the verbal reinforcement of their occurrence; and the use of emotional words in a story-telling task was found to be related also to set and awareness (Ekman, Krasner, & Ullmann, 1963).

The issue of the generalization of a reinforced verbal response has attracted some research attention, but as Greenspoon (1962) notes in a review of such studies, the results have not been conclusive. There is increasing evidence, however, that such an effect does occur, and that it follows principles similar to those underlying the conditioning of motor behavior.

Simkins (1961) found that generalization of a hostile verb response is related both to stimulus similarity and to type of reinforcer. His subjects were matched on the basis of number of hostile verbs and extrapunitive responses scored on one half of the Rosenzweig Picture-Frustration Study. In the first experiment, subjects were positively reinforced for use of hostile verbs in sentences; in the second, they were rewarded with points valued at a penny a point. Following conditioning, all subjects showed generalization to a task consisting of making up sentences using similar verbs; however, only those rewarded with points and pennies

gave an increase in extrapunitive responses on the second half of the Rosenzweig test.

Spreen (1961) and Fadigan (1961) both investigated the problem of whether an increase in the frequency of a response will occur in a subject who is not reinforced, but who is present in a situation in which another subject is being reinforced. These experiments are similar in that a Taffel-type procedure was followed, Spreen using verbs and Fadigan pronouns as the critical response class. Nonreinforced members in Spreen's pairs did not increase their use of the response; Fadigan's did, however.

In a more clinically oriented study, Rogers (1960) hypothesized that frequency of self-reference statements can be changed by reinforcement, and that this reinforcement can alter the self-concept as measured by personality tests. According to reinforcement theory, reported changes of this sort occur through unintentional selective reinforcement by, for example, a Rogerian therapist. In this study, subjects were asked to describe themselves spontaneously in a series of brief interviews. Positive self-references were reinforced for one experimental group and negative for a second. Before and after the interviews, the Manifest Anxiety scale and the *Q* sort Emotional Adjustment Test were administered. No significant differences were found in the post-testing, although improvement was noted in the *Q* sort. Only the negative group showed a conditioning effect. Rogers concluded that the influence of minimal reinforcement is confined to the interview, and that the results support the view that psychotherapy is a process in which the patient learns to talk differently and little else.

The studies using group-therapy patients have attempted to assay the effects of verbal conditioning on later group behavior (Ullmann, Krasner, & Collins, 1961; Ullmann, Krasner, & Ekman, 1961). Both report that groups who were verbally reinforced for the use of emotionally toned words in story-telling sessions gained significantly on ratings of adequacy in interpersonal relationships following the conditioning procedure. This result was felt to reflect a form of role retraining in which the subjects learned that spontaneous expression was appropriate and rewarding.

The application of studies on generalization of acquired verbal responses is obviously most relevant to psychotherapy, though the selection and measurement of a criterion for generalization is a thorny problem. While research so far has dealt with rather simple tasks to study these effects, it is reasonable to assume that relevant variables discovered in the laboratory will be applicable to more complex situations (Miller, 1951) and that eventually clinical practice with a goal of behavior modification will come under the egis of science.

REFERENCES

Adams, J. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, 54, 383–405.

Auld, F., & Murray, E. M. Content analysis of psychotherapy. *Psychol. Bull.*, 1955, 52, 377–395.

Auld, F., & White, A. M. Rules for dividing interviews into sentences. *J. Psychol.*, 1956, 42, 273–281.

Babladelis, Georgia. Personality and verbal conditioning effects. *J. abnorm. soc. Psychol.*, 1961, 62, 41–43.

Binder, A., McConnell, D., & Sjoholm, N. A. Verbal conditioning as a function of experimenter characteristics. *J. abnorm. soc. Psychol.*, 1957, 55, 309–314.

Brown, F., & Webb, W. B. A study of set effects in the verbal conditioning experiment. *Psychol. Rep.*, 1960, 6, 50.

Campbell, J. M., Jr. Verbal conditioning as a function of the personality characteristics of experimenters and subjects. Un-

published doctoral dissertation, University of Washington, 1960.

CROWNE, D. P., & STRICKLAND, BONNIE R. The conditioning of verbal behavior as a function of the need for social approval. *J. abnorm. soc. Psychol.*, 1961, **63**, 395–401.

DAS, J. P. Conditioning and hypnosis. *J. exp. Psychol.*, 1958, **56**, 110–113.

DAS, J. P. Some correlates of verbal conditioning. *Psychol. Stud., Mysore*, 1961, **6**, 30–35.

DAS, J. P., & MITRA, A. K. Relative effectiveness of electric shock and praise and reproof in verbal conditioning. *J. gen. Psychol.*, 1962, **67**, 141–146.

DULANEY, D. E. Hypotheses and habits in verbal operant conditioning. *J. abnorm. soc. Psychol.*, 1961, **63**, 251–263.

EKMAN, P., KRASNER, L., & ULLMANN, L. P. Interaction of set and awareness as determinants of response to verbal conditioning. *J. abnorm. soc. Psychol.*, 1963, **66**, 387–389.

EVERSTINE, L., & BENDIG, A. N. Conditioning neurotic verbalizations. *Amer. Psychologist*, 1960, **15**, 430. (Abstract)

FADIGAN, J. Indirect verbal conditioning. Unpublished master's thesis, Florida State University, 1961.

FERGUSON, D. C., & BUSS, A. H. Operant conditioning of hostile verbs in relation to experimenter and subject characteristics. *J. consult. Psychol.*, 1960, **24**, 324–327.

GEWIRTZ, J. L., & BAER, D. M. Deprivation and satiation of social reinforcers as drive conditioners. *J. abnorm. soc. Psychol.*, 1958, **57**, 165–172.

GREENSPOON, J. The effect of verbal and nonverbal stimuli on the frequency of members of two verbal response classes. Unpublished doctoral dissertation, Indiana University, 1951.

GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.

GREENSPOON, J. Verbal conditioning and clinical psychology. In A. J. Bachrach (Ed.), *Experimental foundations of clinical psychology*. New York: Basic Books, 1962.

GROSS, L. E. Effects of verbal and non-verbal reinforcement in the Rorschach. *J. consult. Psychol.*, 1959, **23**, 66–68.

HAAS, K. Verbal conditioning of affective responses. *J. gen. Psychol.*, 1962, **67**, 319–322.

HALL, W. E. Effects of set and reinforcement in verbal conditioning. *J. gen. Psychol.*, 1960, **63**, 239–248.

HETRICK, W. R., & HAAS, K. Some personality correlates of verbal conditioning. *J. Psychol.*, 1962, **53**, 409–415.

HUMPHREYS, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

JOHNSON, W. Studies in language behavior. *Psychol. Monogr.*, 1944, **56**(2, Whole No. 255).

KANFER, F. H., & KARAS, S. Prior experimenter-subjects interaction and verbal conditioning. *Psychol. Rep.*, 1959, **5**, 343–353.

KANFER, F. H., & MCBREARTY, J. F. Minimal social reinforcement and interview content. *J. clin. Psychol.*, 1962, **18**, 210–215.

KANFER, F. H., & MARSTON, R. Effect of task-relevant information on verbal conditioning. *J. Psychol.*, 1962, **53**, 29–36.

KRASNER, L. The use of generalized reinforcers in psychotherapy research. *Psychol. Rep.*, 1955, **1**, 19–25.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **15**, 148–171.

KRASNER, L. The therapist as a social reinforcement machine. In H. H. Strupp & L. Luborsky (Eds.), *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1962.

KRASNER, L., ULLMANN, L. P., WEISS, R. L., & COLLINS, BEVERLY J. Responsivity to verbal conditioning as a function of three different examiners. *J. clin. Psychol.*, 1961, **17**, 411–415.

KRASNER, L., WEISS, R. L., & ULLMANN, L. P. Responsivity to verbal conditioning as a function of awareness. *Psychol. Rep.*, 1961, **8**, 523–538.

LEVIN, S. M. Effects of awareness on verbal conditioning. *J. exp. Psychol.*, 1961, **61**, 67–75.

MCDONNELL, CAROL R., & INGLIS, J. Verbal conditioning and personality. *Psychol. Rep.*, 1962, **10**, 374.

MCNAIR, D. M. Reinforcement of verbal behavior. *J. exp. Psychol.*, 1957, **53**, 40–46.

MANDLER, G., & KAPLAN, W. K. Subjective evaluation and reinforcing effect of a verbal stimulus. *Science*, 1956, **124**, 582–583.

MILLER, J. Objective methods of evaluating process and outcome in psychotherapy. *Amer. J. Psychiat.*, 1951, **108**, 258–263.

MOCK, J. F. The influence of verbal and behavioral cues of a listener on the verbal productions of the speaker. Unpublished doctoral dissertation, University of Kentucky, 1957.

NUTHMANN, ANNE M. Conditioning of a response class on a personality test. *J. abnorm. soc. Psychol.*, 1957, **54**, 19–23.

POSTMAN, L., & SASSENRATH, J. The automatic action of verbal rewards and punishments. *J. gen. Psychol.*, 1961, **65**, 109–136.

RAZRAN, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–365.

REIDY, MARY E. *A study of the unconscious effects of approval and disapproval on verbal behavior.* Washington, D. C.: Catholic Univer. Press, 1958.

ROGERS, C. R., & SKINNER, B. F. Some issues concerning the control of human behavior: A symposium. *Science*, 1956, **124**, 1057–1066.

ROGERS, J. M. Operant conditioning in a quasi-therapy setting. *J. abnorm. soc. Psychol.*, 1960, **60**, 247–252.

SALZINGER, K. Experimental manipulation of verbal behavior: A review. *J. gen. Psychol.*, 1959, **61**, 56–94.

SAPOLSKY, A. Effect of interpersonal relationships on verbal conditioning. *J. abnorm. soc. Psychol.*, 1960, **60**, 241–246.

SARASON, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *J. abnorm. soc. Psychol.*, 1958, **56**, 339–344.

SCOTT, W. A. Attitude change through reward of verbal behavior. *J. abnorm. soc. Psychol.*, 1957, **55**, 72–75.

SENKO, M. G., CHAMP, R. A., & CAPALDI, E. J. Supplementary report: Resistance to extinction of a verbal response as a function of the number of acquisition trials. *J. exp. Psychol.*, 1961, **61**, 350–351.

SIDOWSKI, J. B. Influence of awareness of reinforcement on verbal conditioning. *J. exp. Psychol.*, 1954, **48**, 355–360.

SIMKINS, L. D. Generalization effects of hostile verb reinforcement as a function of stimulus similarity and type of reinforcer. *J. Pers.*, 1961, **29**, 64–72.

SIMKINS, L. D. Instructions as discriminative stimuli in verbal conditioning and awareness. *J. abnorm. soc. Psychol.*, 1963, **66**, 213–219.

SKINNER, B. F. The generic nature of the concept of stimulus and response. *J. gen. Psychol.*, 1935, **12**, 40–63.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.

SOUTHWELL, E. A. A note on awareness in verbal conditioning. *Psychol. Rep.*, 1962, **10**, 40.

SPIELBERGER, C. D. Awareness in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham, N. C.: Duke Univer. Press, 1962.

SPIELBERGER, C. D., & DENIKE, L. D. Operant conditioning of plural nouns: A failure to replicate the Greenspoon effect. *Psychol. Rep.*, 1962, **11**, 355–366.

SPIELBERGER, C. D., LEVIN, S. M., & SHEPHERD, MARY. The effects of awareness and attitude toward the reinforcement on the operant conditioning of verbal behavior. *J. Pers.*, 1962, **30**, 106–121.

SPREEN, GEORGIA W. Elicitation of imitative responses in a verbal conditioning experiment. Unpublished doctoral dissertation, Indiana University, 1961.

STAATS, A. W. Verbal habit-families, concepts, and the operant level of word classes. *Psychol. Rev.*, 1961, **68**, 190–204.

STAATS, A. W., STAATS, C. K., HEARD, W. G., & FINLEY, J. R. Operant conditioning of factor analytic personality traits. *J. gen. Psychol.*, 1962, **66**, 101–114.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

THORNDIKE, E. L. *Psychology of wants, interests, and attitudes.* New York: Appleton Century, 1935.

ULLMANN, L. P., KRASNER, L., & COLLINS, BEVERLY J. Modification of behavior through verbal conditioning. *J. abnorm. soc. Psychol.*, 1961, **62**, 128–132.

ULLMANN, L., KRASNER, L., & EKMAN, P. Verbal conditioning of emotional words: Effects on behavior in group therapy. *Res. Rep. Veterans Admin. Palo Alto*, 1961, No. 15.

VERPLANCK, W. S. The control of the content of conversation: Reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668–676.

WEISS, R. L. The influence of "set for speed" on learning without awareness. *Amer. J. Psychol.*, 1955, **68**, 425–431.

WEISS, R. L., & ULLMANN, L. P. Responsivity to verbal conditioning as a function of emotional atmosphere and pattern of reinforcement. *Psychol. Rep.*, 1960, **6**, 415–426.

WEISS, R. L., ULLMANN, L. P., & KRASNER, L. On the relationship between hypnotizability and response to verbal operant conditioning. *Psychol. Rep.*, 1960, **6**, 59–60.

WICKES, T. A. Examiner influence in a testing situation. *J. consult. Psychol.*, 1956, **20**, 23–25.

(Received January 10, 1964)

Psychological Bulletin
1964, Vol. 62, No. 6, 394-400

SPONTANEOUS ALTERNATION BEHAVIOR IN HUMANS:

IMPLICATIONS FOR PSYCHOLOGICAL RESEARCH

DUANE P. SCHULTZ

Mary Washington College of the University of Virginia

It is suggested that 1 response-influencing variable in psychological experiments may be the individual S's need for sensory variation as manifested in alternation behavior. The proposition is put forth that in an experimental situation where a discriminatory response to one of a series of successively presented stimuli is required, the organism will seek to optimize the amount of stimulus variability or complexity by alternation behavior. The literature on alternation behavior in human Ss is reviewed and the conditions serving to facilitate alternation are presented. The implications of these facilitating conditions for studies dealing with vigilance behavior, retroactive interference, opinion change, and psychophysics are discussed.

A recent article by Orne (1962) gave further support to the proposition that the human subject in a psychological experiment is not an automatonlike passive responder to stimuli. Orne's focus is not on what is done to the subject but rather on what the human subject does in the experimental situation. He suggests that the subject must be recognized as an active participant whose performance in the experiment is at least partially determined by his own perception of the purpose and nature of the experiment, of the person of the experimenter, of the setting of the laboratory, etc.

The purpose of this paper is to suggest the operation of another response-influencing variable; one which is not related to the social-psychological context of the experimental situation but rather is a function of the subject's individual need for sensory variation as manifested in spontaneous alternation behavior. It is important to note that this point of view while further suggesting that the human subject is not merely a passive responder to stimuli, is more bound to the nature, number, temporal sequence, and manner of presentation of stimuli than Orne's formulation would appear to be, and hence subject to more control by the experimenter.

Fiske and Maddi (1961) define variation in stimulation as, "the extent to which stimulation at a particular moment differs from that which preceded it, or to the average degree of such moment-to-moment changes [p. 12]." Thus, our concern is with the relative heterogeneity of stimulation and the effect on behavior of two or more successive stimulus impacts in a more or less homogeneous sensory environment.

Thus, the proposition is put forth that in an experimental situation where a discriminatory response to one of a series of successively presented stimuli is required, the organism will seek to optimize the amount of stimulus variability or complexity by alternation behavior. Both motor responses in rats and motor and verbal responses in humans show a tendency toward nonrepetition of the preceding or very recent responses. Fiske and Rice (1955) note that:

> It appears likely that the organism does not seek to avoid making the previous response

> but rather seeks to respond in such a way as to vary the total pattern of stimulation reaching it, including the stimulus produced directly or indirectly by its own response [p. 242].

The literature in this area is an outgrowth of Hebb's (1955) thesis in which he ascribed to stimuli the dual function of providing a cue in controlling goal responses and facilitating arousal or vigilance. It was the latter role to which Hebb ascribed motivational properties. Scott (1957) argued that more emphasis should be given to this arousal function of stimuli. Much evidence, both neurophysiological and behavioral, supports the importance of this secondary role for stimuli.

Impulses from the same sensory stimulus have been shown to reach the cerebral cortex via two separate pathways producing both specific and nonspecific effects on behavioral organization. The specific effect is produced by impulse passage along the sensory tract through the corresponding nucleus in the thalamus and terminating on a specific projection area of the cortex. The nonspecific effects are produced by the slower passage of impulses (from the same stimuli) through the ascending reticular activating system and terminating in diffuse bombardment over wide areas of the cerebral cortex. It is this latter type of nonspecific stimulation that is considered necessary for attaining and maintaining alert behavior. This literature has been surveyed by Hebb (1955), Scott (1957), Lindsley (1957), Malmo (1959), and Samuels (1959).

What is the effect on behavior of this diffuse bombardment of areas of the cerebral cortex? Consider a situation with a minimum (or homogeneous level) of variation in stimulation such as a series of repetitions of the same or similar stimuli. Adaptation to such a repetitive situation readily develops and alertness will decrease. If a novel, that is, new and different, stimulus is introduced, either through experimental manipulation or through the subject changing his response behavior, this variation will immediately increase alertness and the vigor of ongoing behavior (Fiske & Maddi, 1961), with its concomitant focusing of attention (cognitive change) and arousal of interest (motivational change). Scott (1957) points out that the effect of change in the stimulus environment not only focuses attention and interest on the novel stimulus, but it also has the capability of altering the general character of the individual's relation with the environment. This is seen in the functioning of the selective processes.

What happens when these novel stimuli are repeatedly presented to the organism? Scott (1957) suggests that with continued exposure the stimuli will lose their nonspecific effects. This process, termed "sensory habituation" results in attention being no longer focused on the (once) novel stimuli. Hence, we would expect performance, that is, overt response to the continuing stimuli, to deteriorate until a new stimulus is introduced.

The literature dealing with alternation behavior in rats has been surveyed elsewhere (Dember, 1960; Dember & Earl, 1957; Dember & Fowler, 1958; Fiske & Maddi, 1961; Glanzer, 1953, 1958; Walker, 1958) and need not be reviewed here beyond pointing out that in alternating, the rat would seem to be seeking stimulation which differs from that to which it was previously exposed.

There are also a number of studies to be reviewed below, which show that humans demonstrate alternation behavior in experimental situations where the trials are contiguous or close in time and where no strong, specific motiva-

tions are involved, that is, no one stimulus-response category is reinforced.

Studies of Spontaneous Alternation

The variability of a human subject's responses in repetitions of the same motor task was noted by Thorndike (1923).

Telford (1931) questioned whether an effect similar to the refractory phase in sensitive tissue could be observed in voluntary and associative responses. He concluded that in the three processes studied (reaction time, judging the longer of two parallel lines, and nonsense syllable-number sequences), effects were produced in the organism which served as a barrier against immediate repetition.

Alternation of response was also noted by Wingfield (1943). Using a variety of manual alternation situations, he concluded that: forcing an increase in the speed of reaction reduces the number of spontaneous alternations; increasing the difference between the alternatives gives a corresponding increase in alternations; and if the previous choices are successful, there is a decreased likelihood of alternation.

Investigating Hull's principle of reactive inhibition, Siegel (1950) found that the frequency of occurrence of one response increased rectilinearly with the amount of prior exercise in the execution of the other response. He used two light switches placed 30 inches apart and, first, gave varying amounts of exercise in turning off the switch located on the right, followed by trials in which either the right or the left switch could be used.

In exploring relative response frequencies, Shelly (1958) associated conditional probabilities of responses leading to reinforcement, with preceding responses. These conditional probabilities were such as to never (except in a control condition) lead to a reinforcement for the repetition of a response. The results showed that the subjects tended to change alternatives on successive trials more frequently than would occur by chance.

Sugimura and Iwahara (1958) investigated alternation by requiring subjects to make 21 consecutive guesses, with no information about their success as to whether the correct response was right or left. As had been predicted, the percent of alternation decreased significantly with the increase in intertrial interval. The effect increased when the two test objects were dissimilar.

Using paper and pencil mazes with human subjects, Denny and Allen (1959) had subjects traverse the inverted L maze 10 times after which they were presented with a single T maze of like dimensions at intervals of 0, 24, 48, 72, or 96 hours. The amount of alternation for each of the time-delay intervals was 90%, 80%, 67%, 50%, and 50%, respectively. After 3 days the satiation effects were dissipated. The results also showed that doubling the length of the arm increased the satiation effect significantly and group administration of the task decreased the effect significantly.

Iwahara (1959) reported two studies on spontaneous alternation. In Experiment I, subjects were asked to press either the right or the left key. A red light flashed when the right key was pressed and a green light flashed at the pressing of the left key. The subjects were told that one of the two responses was correct but no information was given them of their results. Spontaneous choice alternation was found to decrease slightly from 0- to 30-second intertrial intervals for children but no such tendency was shown for adults. However, his results showed considerable decrease

in alternation behavior with greater intertrial intervals for both age groups, and adults indicated consistently less alternation than children. Experiment II tested the effect of similarity of choice objects on response alternation. Alternation tendency was found to be greater when the two stimulus objects were differently colored than when they were similarly colored.

Using the human stylus maze, Lawless and Engstrand (1960) tested alternation and found that the distance between a forced 90-degree turn and the choice point was a relevant variable but that time between turns was not.

Alternation in Psychophysical Experiments

The tendency to avoid the repetition of the preceding judgment using the psychophysical method of constant stimuli was noted by Fernberger as long ago as 1920. Similar results were noted by Turner (1931), Arons and Irwin (1932), and Irwin and Preston (1937). Turner (1931) found that when a standard weight was preceded by a heavier weight with which it was not compared, a comparison weight following that standard weight was judged "lighter" a greater number of times. If the preceding weight was lighter than the standard the comparison weight following the standard was judged "heavier" a greater number of times.

Arons and Irwin (1932) obtained psychophysical judgments on objectively equal standard and comparison stimuli and found that five out of seven subjects avoided repetition of judgments in the same category. Also using the method of constant stimuli, Irwin and Preston (1937) showed that judgments on material presented in one sense modality predisposed their subjects to avoid repetition of the same category in making judgments on material presented in another modality. Thus, they concluded that the tendency to avoid repetition of judgments cannot be attributed to changes in the sensory processes themselves. Also, having required their subjects to give alternating comparisons in the two modalities, stipulating that the report in the first modality be given in one set of terms and the report in the second modality be given in a second set of terms, they showed that the avoidance tendency was not dependent upon the mechanisms involved in the giving of the response words themselves.

The nonindependence of successive responses in measuring visual threshold has been reported by Verplanck, Collier, and Cotton (1952) and by Wertheimer (1953). The results of the study by Verplanck et al. showed that when the subject is presented with a discrete dim flash of light, his response to it depended not only upon the luminance of the flash, but also on how he responded to preceding flashes of light of the same luminance. Wertheimer (1953) obtained measurements of auditory, visual, and pain thresholds at intervals of 6 seconds, 1 minute, 3 minutes, and 1 day. His results indicated that the threshold variations in time were not random.

Thus, the literature suggests four conditions which can serve to facilitate alternation behavior in humans: (*a*) where there is no reinforcement and/or knowledge of correctness of response (Fiske & Rice, 1955), (*b*) where there is a greater dissimilarity between stimuli (Iwahara, 1959; Sugimura & Iwahara, 1958; Wingfield, 1943), (*c*) the greater the prior exercise of one response alternative (Denny & Allen, 1959; Siegel, 1950), (*d*) a short intertrial interval (Iwahara, 1959; Sugimura & Iwahara, 1958).

Discussion and Implications

Many experimental paradigms fulfill one or more (or all) of the above conditions. Consider the situation in which the subject is presented with a series of similar stimulus events in which his task is to detect change over relatively long periods of sustained observation. Such is the case with the study of human vigilance behavior. Scott (1957) suggested that the loss of efficiency over time in vigilance tasks was directly related to a reduction in stimulus variation. He noted that one would expect to find performance restored to or maintained at a higher level under conditions which increased the variety of either peripheral or task stimuli, for example, rest periods, high signal rate, knowledge of results, interpolated messages, multiple-stimulus sources. Frankmann and Adams (1962) suggest that it would be fruitful to look for an expression of the stimulus variation hypothesis in terms of stimulus control of responses, both environmentally produced and response produced.

There is also the experimental situation in which the subject is presented with two or more stimuli in temporal succession (in some cases with a time and/or task interpolation between the stimulus presentations) and is asked to make some overt response (immediately or after a time delay) to one of the stimuli. Such would be the case, for example, with the retroactive interference paradigm, with certain of the psychophysical methods, and with studies involving opinion change.

As an example, let us consider the literature dealing with primacy-recency conditions of opinion change. Primacy-recency refers to that communication situation in which two opposing positions of a single issue are presented and endorsement of or agreement with one of the positions is called for. Primacy refers to the case where the communication presented first is more effective in inducing opinion change in the direction of its argument and recency refers to the case where the second argument is so effective. Schultz (1963) suggested an interpretation of the primacy-recency data within a sensory variation framework and noted that,

> with the stimuli in the second trial being similar to those of the first, the experimental situation provides a minimal, homogeneous level of variation. Hence, in order to maintain the level of activation produced by the first stimulus presentation the individual will tend to retain focus on this first stimulus [pp. 130–131].

In the situation where an interpolated response as well as a response after the second presentation are called for, Schultz suggests that this alternation would manifest itself in giving responses which differ from those given previously. He further suggests that when there is a time and/or task interpolation between the stimulus presentations one would expect the capacity to respond (increase in focused attention) to reappear spontaneously since the delayed presentation of the stimulus would again contribute to sensory variation, that is, the stimulus now differs from the preceding one.

Working within this sensory-variation framework, Schultz attempted to interpret the primacy-recency data in terms of the organism's quest for an optimal level of variation in stimulation rather than in terms of strictly attitudinal phenomena.

It is suggested that data on other varieties of human behavioral functions are also amenable to systematic interpretation within this same framework.

As an example let us again consider psychophysics. Wertheimer (1953) noted that since the inception of psy-

chophysics, notice has been taken of the fact that successive measurements of a threshold are not identical. This variation had been generally attributed to experimental error and, accordingly, thresholds have been defined in statistical terms on the assumption that the human subject is stable. More recently, interest has arisen in the temporal course of threshold measurements from the point of view that the subject may be unstable. The data discussed earlier (Arons & Irwin, 1932; Fernberger, 1920; Irwin & Preston, 1937; Turner, 1931; Verplanck, Collier, & Cotton, 1952; Wertheimer, 1953) indicate that the threshold variations in time are not random.

Findings such as these, when interpreted in terms of spontaneous alternation behavior, would seem to have ramifications wherever comparable measurements are made, for, as Smith (1951) suggests, successive responses depend more or less on preceding stimulus-response situations. The human subject responds not only to the represented stimulus, but also to the previous presentation and/or to his previous response.

Shelly (1958) suggests that,

> situations in which the relative frequencies of events are dependent upon preceding responses might be considered to be an elementary prototype of social situations. One of the most characteristic features of social behavior is that the consequences of a person's behavior depend partly on preceding events which consist largely of his own responses [p. 239].

There would seem to be no valid reason for assuming that the human organism sheds this behavior when he becomes a subject in a psychological experiment.

It is only with an increased understanding of such factors as presented in this paper and that of Orne (1962), that the experimental method of psychology can become a more effective tool in predicting behavior.

REFERENCES

Arons, L., & Irwin, F. W. Equal weights and psychophysical judgments. *J. exp. Psychol.*, 1932, **15**, 733–756.

Dember, W. N. *The psychology of perception.* New York: Holt, 1960.

Dember, W. N., & Earl, R. W. Analysis of exploratory, manipulatory, and curiosity behavior. *Psychol. Rev.*, 1957, **64**, 91–96.

Dember, W. N., & Fowler, H. Spontaneous alternation behavior. *Psychol. Bull.*, 1958, **55**, 412–428.

Denny, M. R., & Allen, J. N. Alternation behavior in humans. *Amer. Psychologist*, 1959, **14**, 405. (Abstract)

Fernberger, S. W. Interdependence of judgments within the series for the method of constant stimuli. *J. exp. Psychol.*, 1920, **3**, 126–150.

Fiske, D. W., & Maddi, S. R. *Functions of varied experience.* Homewood, Ill.: Dorsey Press, 1961.

Fiske, D. W., & Rice, Laura. Intra-individual response variability. *Psychol. Bull.*, 1955, **52**, 217–250.

Frankmann, Judith, & Adams, J. A. Theories of vigilance. *Psychol. Bull.*, 1962, **59**, 257–272.

Glanzer, M. The role of stimulus satiation in spontaneous alternation. *J. exp. Psychol.*, 1953, **45**, 387–393.

Glanzer, M. Curiosity, exploratory drive, and stimulus satiation. *Psychol. Bull.*, 1958, **55**, 302–315.

Hebb, D. O. Drives and the CNS (conceptual nervous system). *Psychol. Rev.*, 1955, **62**, 243–254.

Irwin, F. W., & Preston, M. G. Avoidance of repetition of judgments across sense modalities. *J. exp. Psychol.*, 1937, **21**, 511–520.

Iwahara, S. Studies in spontaneous alternation in human subjects: III. A developmental study. *Jap. psychol. Res.*, 1959, **8**, 1–8.

Lawless, R. H., & Engstrand, R. D. Alternation in the human stylus maze. *Psychol. Rec.*, 1960, **10**, 101–106.

Lindsley, D. B. Psychophysiology and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957.* Lincoln: Univer. Nebraska Press, 1957. Pp. 44–105.

Malmo, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

Orne, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Amer. Psychologist,* 1962, **17**, 776–783.

Samuels, Ina. Reticular mechanisms and behavior. *Psychol. Bull.,* 1959, **56**, 1–25.

Schultz, D. P. Primacy-recency within a sensory variation framework. *Psychol. Rec.,* 1963, **13**, 129–139.

Scott, T. N. Literature review of the intellectual effects of perceptual isolation. Report No. HR66, July 1957, Department of National Defense, Defense Research Board, Canada.

Shelly, M. W. The effects of response contingent probabilities which favor response change. *J. exp. Psychol.,* 1958, **56**, 239–245.

Siegel, P. S. Reactive inhibition as a function of number of response evocations. *J. exp. Psychol.,* 1950, **40**, 604–608.

Smith, B. B. On some difficulties encountered in the use of factorial designs and analysis of variance with psychological experiments. *Brit. J. Psychol.,* 1951, **42**, 250–268.

Sugimura, T., & Iwahara, S. Studies of spontaneous alternation in human subjects: I. Effects of inter-response intervals and heterogeneity of stimulus objects on response alternation in children. *Jap. J. Psychol.,* 1958, **28**, 358–365.

Telford, C. W. The refractory phase of voluntary and associative responses. *J. exp. Psychol.,* 1931, **14**, 1–36.

Thorndike, E. L. The variability of an individual in repetitions of the same task. *J. exp. Psychol.,* 1923, **6**, 161–167.

Turner, W. O. Intra-serial effects with lifted weights. *Amer. J. Psychol.,* 1931, **43**, 1–25.

Verplanck, W. S., Collier, G. H., & Cotton, J. W. Nonindependence of successive responses in measurements of the visual threshold. *J. exp. Psychol.,* 1952, **44**, 273–282.

Walker, E. L. Action decrement and its relation to learning. *Psychol. Rev.,* 1958, **65**, 129–142.

Wertheimer, M. An investigation of the "randomness" of threshold measurements. *J. exp. Psychol.,* 1953, **45**, 294–303.

Wingfield, R. C. Some factors influencing spontaneous alternation in human subjects. *J. comp. physiol. Psychol.,* 1943, **35**, 237–243.

(Received February 3, 1964)

Psychological Bulletin
1964, Vol. 62, No. 6, 401–410

AMPHETAMINE TOXICITY, POPULATION DENSITY, AND BEHAVIOR:

A REVIEW [1]

D. D. THIESSEN

Scripps Clinic and Research Foundation, La Jolla, California

Administration of the drug amphetamine to isolated mice heightens behavioral arousal, increases adrenocortical and catecholamine activity, raises body temperature, stimulates neurological activity, and can cause death. All of these effects are exaggerated when mice are grouped, suggesting that social interaction potentiates those processes common to amphetamine. A review of the effects of amphetamine on grouped mice is presented along with an attempt to link the physiological mechanisms of amphetamine toxicity to those activated by any increase in population size.

For several years it has been known that aggregating mice within a confined space increases behavioral excitement and potentiates the lethality of amphetamine (Chance, 1946; Gunn & Gurd, 1940). It has since been found that these effects are linked with body temperature, adrenal activity, and depletion of body reservoirs of catecholamines and are modifiable by difference in group size, ambient temperature, genotype, and past experience. Presumably, the novelty of grouping, aggression, and heightened excitement act to mobilize physiological actions that add to the stimulating properties of amphetamine. In combination their effects can be excessive, physiological mechanisms are overtaxed and death results. In this review an attempt is made to present the basic relations between grouping and amphetamine toxicity and describe the known behavioral and physiological mediators. Nearly all experimental effort has involved the mouse (*Mus musculus*). Of all density responses investigated (for partial reviews, see Thiessen, 1964a; Thiessen & Rodgers, 1961) the mechanisms of amphetamine toxity are perhaps best elucidated.

The Drug

Amphetamine is described as a synthetic sympathomimetic (adrenergic) amine having a potent stimulating effect on the central nervous system (Grollman, 1960). Its primary site of action is at the lower brain centers, including the midbrain reticular system, hypothalamus, and limbic structures (Berger, 1960; Sharp, Nielson, & Porter, 1962). Barbiturates, tranquilizers, and other central nervous system depressants effectively antagonize the stimulating actions of amphetamine (Grollman, 1960; Lasagna & McCann, 1957). Amphetamine is marketed as a racemic amphetamine phosphate or sulfate (*dl*-amphetamine or Benzadrine) and as dextroamphetamine sulfate (*d*-amphetamine or Dexidrine), both with the empirical formula of $(C_9H_{13}N)_2 \cdot H_2SO_4$. The dextrorotatory isomer of amphetamine sulfate (*d*) exerts a greater stimulating effect on the central nervous system than does the racemic compound,

[1] This research was supported in part by Grant HM-06139 from the National Institute of Mental Health. Appreciation is extended to David A. Rodgers for helpful comments on the manuscript.

whereas the racemic variety predominates in cardiovascular activity (Grollman, 1960; Moore, 1963). The similarity of amphetamine to endogenously produced hormones (e.g., epinephrine and norepinephrine) and its extensive use as a pharmacologic "psychic energizer" have led to numerous reports on its action. The effects are multiple. Amphetamine has been reported to increase general activity (Gunn & Gurd, 1940), alter conditioning processes (Hunt & Beckwith, 1956), depress food intake (Sharp et al., 1962), initiate tonic-clonic convulsions (Swinyard, Clark, Miyahara, & Wolf, 1961), cause death (Warren & Werner, 1946), raise body and brain temperatures (Askew, 1962; Rothballer, 1959), modify cardiovascular responses (Moore, 1963), and deplete body stores of epinephrine and norepinephrine (McLean & McCartney, 1961). Many of these effects are exaggerated by population density and thus form the basis for explaining amphetamine toxicity under conditions of social interaction.

The Effect

Gunn and Gurd (1940) were first to describe the excitement that occurred when white mice [2] were grouped six to a cage following injection of *dl*-amphetamine and Chance (1946), in turn, noted that the toxicity of this drug was markedly enhanced when mice were grouped. Both of these observations have been confirmed repeatedly (Askew, 1962; Burn & Hobbs, 1958; D'Arcy & Spurling, 1961; Fink & Larson, 1962; Greenblatt & Osterberg, 1961; Hardinge & Peterson, 1963; Höhn & Lasagna, 1960; Lasagna & McCann, 1957; Moore, 1963; Swinyard et al., 1961; Weaver & Kerley, 1962). Amphetamine initiates running, startle responses, aggression, and squealing (Askew, 1962; Chance, 1946, 1947; Greenblatt & Osterberg, 1961; Höhn & Lasagna, 1960; Moore, 1963). Importantly, perhaps, heightened excitement is not ordinarily noted in isolated animals given the drug nor in aggregated mice without the drug, suggesting that amphetamine sensitizes an animal to social stimuli (however, see Moore, 1963; Thiessen, 1964b). At high enough dose levels and especially for isolated animals tonic-clonic convulsions appear within the first few minutes and death quickly follows (Askew, 1962; Höhn & Lasagna, 1960; Moore, 1963). If this sequence of events does not occur, death may still result but at a later time (usually within 4–7 hours) and the proximal symptom appears to be exhaustion (Moore, 1963; Swinyard et al., 1961). Grouped animals in particular die following a period of lethargy and coma, suggesting that isolated and grouped mice succumb to amphetamine for different reasons.

As compared to isolated mice amphetamine toxicity is potentiated in group sizes of 3 (Hardinge & Peterson, 1963; Höhn & Lasagna, 1960; Lasagna & McCann, 1957; Swinyard et al., 1961), 4 (Moore, 1963), 5 (Askew, 1962; D'Arcy, 1962; Höhn & Lasagna, 1960), 8 (Chance, 1946; D'Arcy & Spurling, 1961), 9 (Hardinge & Peterson, 1963), 10 (Burn & Hobbs, 1958; Chance, 1946; Fink & Larson, 1962; Greenblatt & Osterberg, 1961; Weaver & Kerley, 1962), 16, and 32 (Chance, 1946). Moreover, a positive relation has been found between group size and amphetamine mortality (Chance, 1946), although this relation can vary with changes in genotype (Weaver & Kerley, 1962), environmental temperature (Hardinge & Peterson, 1963), and probably other unknown factors. For example, Chance (1946) did not find in-

[2] Where possible strain nomenclature is presented.

creases in drug toxicity for CFF mice under densities of 8 whereas other investigators have, working with different strains. Höhn and Lasagna (1960) were unable to find an increase in amphetamine toxicity between populations of 3 and 5 for Swiss albino mice, but Fink and Larson (1962) found an increase in toxicity between populations of 5 and 10 for CF_1 mice. Similarily, Hardinge and Peterson (1963) were unable to detect aggregation effects for Swiss-Webster mice unless the ambient temperature approximated 26° C. Even at this crucial temperature amphetamine mortality was no different between population sizes of 3 and 9. The importance of genetic differences among strains of mice is emphasized by several investigations (Askew, 1962; Chance, 1946; Greenblatt & Osterberg, 1961; Hardinge & Peterson, 1963; Lasagna & McCann, 1957; Moore, 1963; Swinyard et al., 1961; Weaver & Kerley, 1962), when strains are compared at room temperature without equilibration for population size, sex, age, or other randomly varying conditions. Strain differences in interaction between grouping and amphetamine mortality are often striking; yet it appears that all strains tested at room temperature and under communal circumstances respond to amphetamine with increased mortality. Further, as noted above, when increases in density do result in altered sensitivity to amphetamine the direction of response is predictable.

As might be expected several other central nervous system stimulants exaggerate mortality rates when administered to grouped mice. Chance (1946) reported that *dl*-amphetamine, methedrine, ephedrine, and adrenaline are highly toxic for grouped GFF mice. Amphetamine was particularly effective in this regard. Greenblatt and Osterberg (1961) reported that 8 of 13 stimulants tested hasten death in Manor Farms mice grouped 10 to a cage. In order of toxicity these are methamphetamine, *dl*-amphetamine, methylphenidate, *B*-phenylisopropylhydrazine, pipradrol, amiphenazole, caffeine, and picrotoxin. Interestingly these drugs differ widely in chemical structure, site of action, and secondary effects (Grollman, 1960); nevertheless all are significantly more toxic for grouped than for isolated animals. At this time there is little reason to suspect that all effective stimulants are associated with identical physiological processes.

The Mechanism

Several major "sociobiological" changes have been reported to account for amphetamine toxicity: increased body temperature, heightened excitement (behavioral and neurological), enhanced anxiety or fear, increased adrenocortical activity, and depletion of body stores of catecholamines. Increases in body temperature and neurological excitement, however effected, are particularly important and may represent "last common pathways" between behavior and amphetamine mortality.

Temperature

The notion that body temperature regulates amphetamine mortality arises from the existing association between environmental and body temperature as well as from the demonstration that sympathomimetic amines raise body temperature (Hardinge & Peterson, 1963; Warren & Werner, 1946) and that high environmental temperature often leads to enhanced toxicity in isolated or grouped mice (Askew, 1962; Chance, 1946, 1947; Fink & Larson, 1962; Hardinge & Peterson, 1963; Warren & Werner, 1946). Generally, grouped animals are more sensitive to changes

in environmental temperature. Amphetamine LD_{50} for isolated animals housed at environmental temperatures ranging from 16° C. to 32° C. is about 56% lower at the upper temperatures, while LD_{50} for grouped animals at environmental temperatures ranging from 21° C. to 27° C. is about 78% lower at the upper temperatures. Thus even a small increase in temperature for grouped animals exaggerates drug toxicity. Both Fink and Larson (1962) and Hardinge and Peterson (1963) report an interaction between environmental temperature, drug toxicity, and grouping—as temperature increases, toxicity becomes more pronounced for grouped animals. However, if aggregated mice are exposed to a rapidly moving stream of air at a wide range of temperatures, differences between isolated and grouped mice disappear (Askew, 1962; Hardinge & Peterson, 1963). Further, a direct reduction in ambient temperature from 25° C. to 10° C. can give protection to aggregated mice, even if exposure to the lower temperature is only for a few minutes just before or after injection of amphetamine (Höhn & Lasagna, 1960). Presumably, protection from amphetamine results from a greater dissipation or reduction in body heat. While a drop in temperature is protective, should ambient temperature and hence body temperature drop too low, amphetamine again increases in toxicity. Höhn and Lasagna (1960) found a significant increase in mortality in aggregated mice when the ambient temperature was lowered from 25° C. to 7° C. or 3° C. Perhaps a significant deviation from thermoneutrality in either direction potentiates the effects of amphetamine.

Direct measurements of body temperature generally confirm the relation between heat regulation and amphetamine toxicity. Askew (1962) reports that mice housed five to a group respond to amphetamine with a greater rise in rectal temperature than do isolated mice. Further, as environmental temperatures increase amphetamine will raise body temperature above that expected when either high environmental temperature or amphetamine is applied separately. After several tests with aggregated mice, Askew concluded that irrespective of the dose of amphetamine given and regardless of the housing condition and external temperature, all those mice whose rectal temperature remained below 41.7° C. survived while, with a rare exception, all those whose rectal temperatures rose above 42.4° C. died. Isolated mice also succumbed at approximately 42.4° C. but less frequently attained this lethal body temperature. Chlorpromazine and phenoxybenzamine antagonized both rises in body temperature and mortality. Conversely, 4-Methyl-E-(*B*-chlorenthyl)-thiazole and *L*-thyroxine increased both body temperature and amphetamine toxicity. Calcium acetylsalicylate which had no effect on body temperature was also without effect on mortality. In no case did these compounds enable mice to withstand a rise in body temperature that would normally be fatal to untreated mice.

Variations in body temperature cannot, however, account for all deaths due to stimulating drugs. Swinyard et al. (1961) found approximately the same hyperthermic response to amphetamine for isolated and grouped mice even though aggregated mice died more frequently. Correspondingly, Greenblatt and Osterberg (1961) found that toxic drugs could produce either hyper- or hypothermia. Of those drugs that were more toxic to grouped mice methamphetamine, *dl*-amphetamine, methylphenidate, and amiphenazole caused an increase in rectal temperature, whereas

B-phenylisopropylhydrazine, caffeine, and picrotoxin caused a decrease in rectal temperature. It is possible, of course, that the physiological events leading to death are quite drug specific. It is of interest, however, to point out that almost all of the compounds reported to reduce amphetamine toxicity are antagonists of adrenaline (Askew, 1962; Lasagna & McCann, 1957; Maxwell, 1959), itself a hyperthermic agent (Turner, 1961). The close association between environmental temperature, body temperature, and toxicity of at least one drug, amphetamine, suggests a critical mechanism.

Excitement

Changes in behavior resulting from grouping may potentiate the action of amphetamine. A common observation is that grouping increases general activity, aggression, startle reactions, and squealing (Barnett, 1963; Southwick & Bland, 1959; Thiessen, 1963, 1964b; Thiessen, Zolman, & Rodgers, 1962) and that these responses are exaggerated by amphetamine (Chance, 1946; Gunn & Gurd, 1940; Greenblatt & Osterberg, 1961; Höhn & Lasagna, 1960; Maxwell, 1959). It is likely that the synergistic association between behavior and amphetamine toxicity acts through common physiological mechanisms, perhaps involving decreases in thresholds for neural arousal as well as increases in body temperature.

Swinyard et al. (1961) have explored the relation of amphetamine toxicity to housing experience and central nervous system arousal. Mice aggregated three per cage showed the typical susceptibility to amphetamine; however, increasing the degree of experience with group housing prior to the administration of the drug strikingly reduced the mortality rate. D'Arcy (1962) has replicated this finding. Swinyard and his colleagues went on to show that grouping per se lowers pentylenetetrazol (Metrazol) seizure thresholds and is particularly marked when amphetamine is given. Together these results suggest that grouping effects and amphetamine stimulation summate to enhance central neural activity and that this activity is depressed by previous exposure of animals to each other. It may be, as these investigators infer, that drug toxicity is the result of a high level of neural activity that can be reduced in amount by prior adaptation to noxious social stimuli (anxiety and fear-producing stimuli). On the basis of this hypothesis, stressors of various sorts would be expected to enhance amphetamine toxicity and reduce seizure threshold. This has been confirmed in a number of ways. Restricting cage size has been variably related to amphetamine toxicity (Chance, 1947; Fink & Larson, 1962; Hardinge & Peterson, 1963; Höhn & Lasagna, 1960), but in general increases the potency of amphetamine. Body restraint also has the effect of decreasing electroshock and pentylenetetrazol seizure thresholds and may be regarded as mediated by catecholamine secretion from the adrenals (Swinyard, Radhakrishnan, & Goodman, 1962). Foot shocks to isolated mice have been reported to significantly decrease the LD_{50} of amphetamine (Weiss, Laties, & Blanton, 1961) and are to an extent related to increases in body temperature (Askew, 1962). Forced exercise of isolated mice will likewise lower LD_{50} threshold to a level found when animals are grouped (Hardinge & Peterson, 1963). Crowding per se is typically associated with increases in adrenal steroid and catecholamine secretion (Thiessen & Rodgers, 1961) while amphetamine stimulates the depletion of catecholamines from body stores, especially for mice housed as a group (Moore,

1963). Moreover, isolated mice pretreated with cortisol or long-acting corticotropin show a greater frequency of death when given amphetamine (D'Arcy & Spurling, 1961). Corticotropin additionally enhances amphetamine toxicity in grouped animals, suggesting that endogenous adrenal steroids are more effective in this regard than injected cortisol. Overall, it appears that those factors, however diverse, that stimulate adrenocortical or catecholamine secretion act to enhance the lethality of amphetamine. Perhaps the decisive final actions are a rise in body temperature and excessive neurological activity.

Conversely, conditions which antagonize adrenal steroid and catecholamine activity or attenuate behavioral and neurological excitement also reduce amphetamine toxicity. Implicating adrenal activity is the observation by Weiss et al. (1961) that adrenalectomized mice withstand doses of amphetamine that are normally toxic for intact animals. Several drugs which either depress andrenergic activity or reduce behavioral and neurological excitement have a similar effect. Reserpine, chlorpromazine, dibenzyline, dibozane, phenoxybenzamine, phenotolamine, phenylhomoveratryl-piperazine, piperoxane, and promazine, among others already mentioned, reduce amphetamine toxicity within groups of mice and frequently are highly effective at dose levels which depress motor activity (Askew, 1962; Lasagna & McCann, 1957; Maxwell, 1959; Weiss et al., 1961). Interestingly, reserpine can also reduce adrenal hypertropy due to the grouping of mice (Christian, 1956). Maxwell (1959) found a close association between the ability of several of the drugs listed above to reduce amphetamine toxicity in grouped mice and depression of motor activity (rho computed from his data = .86; $p < .05$). For at least two of these agents, reserpine and chlorpromazine, as dose level increases, the number of animals dying from amphetamine decreases (Burn & Hobbs, 1958). This relation is set down in Table 1. The degree of protection afforded by chlorpromazine was particularly marked for grouped animals, suggesting that grouped animals are not only more susceptible to drug lethality but also to conditions that reduce mortality. This has been confirmed for reserpine and a

TABLE 1

EFFECT OF TRANQUILIZERS ON *DL*-AMPHETAMINE TOXICITY IN MICE

Reserpine		Chlorpromazine			
Grouped[a]		Isolated[b]		Grouped[c]	
mg/kg	% mortality	mg/kg	% mortality	mg/kg	% mortality
.00	76	.00	88	.00	93
.06	70	.30	75	.10	77
.10	61	.60	64	.17	67
.17	44	1.20	45	.24	50
.29	27	2.40	15	.60	13

Note.—Burn and Hobbs, 1958.
[a] 20 mg/kg *dl*-amphetamine.
[b] 100 mg/kg *dl*-amphetamine.
[c] 14 mg/kg *dl*-amphetamine.

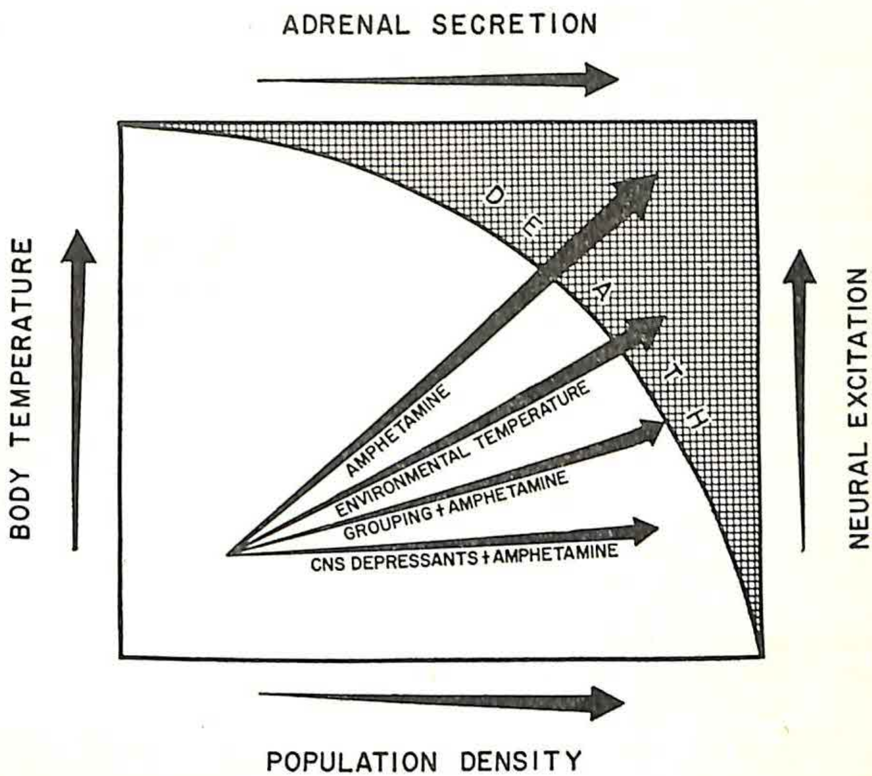


FIG. 1. Relations between population density, physiological response, and amphetamine toxicity. (Explanation in text.)

variety of adrenergic blocking drugs (Maxwell, 1959). Exposure to cold is another means used with crowded mice to depress excitement and hence amphetamine toxicity. Mice exposed to 10° C. for varying periods of time showed less ambulatory behavior and a 50% reduction in amphetamine mortality (Höhn & Lasagna, 1960). All of these effects are consistent with the interpretation that a lowered output of adrenal products, reduction in body temperature, and a depression of behavioral and neurological activity act in opposition to the stimulating effects of amphetamine. Again the final common pathway may involve temperature regulation and neurological excitement.

The Theory

Evidence appears to document a relation between crowding, behavioral excitement, secretory action of the adrenals, heightened neurological activity, increased body temperature, and amphetamine toxicity. Presumably intense social interaction initiates adrenal steroid and catecholamine secretions which arouse central nervous system activity and increase body temperature. Amphetamine, to an extent, stimulates identical pathways. In this sense then, pharmacological activity duplicates and exaggerates those processes already activated by grouping and should not be viewed apart from other physiological processes of population dynamics. The difference lies not in mode of operation but in degree and consequence. Dense populations can collapse without amphetamine (Thiessen & Rodgers, 1961). What amphetamine does, apparently, is to add to the physiological load already balanced at a precarious level—the

TABLE 2

RANK-ORDER CORRELATION OF AMPHETAMINE LETHALITY, MOTOR ACTIVITY, AND RECTAL TEMPERATURE IN ISOLATED AND GROUPED MICE

	Isolated		Grouped	
	Motor activity	Rectal temperature	Motor activity	Rectal temperature
Lethality potency ratio: $\frac{\text{Isolated}}{\text{Grouped}}$ LD_{50}	.31	.39	.65*	.78*

Note.—Greenblatt and Osterberg, 1961.
* $p < .05$.

physiological burden becomes excessive and death results.

Adrenal action, neural excitement, and body temperature must be primary mediating mechanisms. The data examined suggest that any means used to stimulate these systems enhance the toxicity of amphetamine while any means used to depress these systems afford considerable protection from amphetamine. Figure 1 presents a relational diagram that summarizes the interactions assumed to be operating when population density increases. The borders of the figure show the most important relations for any increase in density. The arrows indicate direction of response as well as degree. The clearest relations are between the opposing border arrows—as density increases, adrenal secretion increases and as body temperature rises, so, presumbly, does neural excitation. Relations between adjacent border arrows are probably valid; for example, as adrenal secretion rises so does body temperature and neural excitation. Thus, population density per se is seen to initiate adrenal secretion, a rise in body temperature, and an increase in neural excitation. It is assumed that any of these factors acting singly or in combination can result in death if activated to excess. Superimposed on these relations are those processes associated with amphetamine. These are depicted in the body of the figure by arrows differing in slope and length. As population density increases amphetamine potentiates the rise in adrenal secretion, body temperature, and neural excitation. The threshold for death is quickly reached. An increase in environmental temperature has a similar effect but is slower acting and less likely to end in fatalities. Prior experience with grouping attenuates all the stimulating qualities of amphetamine and prevents death. Central nervous system depressants appear to block amphetamine and population death most effectively.

The above schema is consistent with existing data and has the advantage of incorporating the effects of amphetamine with those common to density per se. Additional effects of stress and other manipulations of amphetamine toxicity can easily be integrated into this broader picture. However, it is not clear what weight should be given to the various factors. Too, while it is assumed that social interaction is the proximal stimulus for the critical series of events, just what type of interaction is involved and how closely it relates to amphetamine toxicity is little known. Greenblatt and Osterberg (1961) do suggest that body temperature is more closely associated with drug toxicity than is motor activity. These investigators related motor activity in isolated

and grouped mice with the maximum body temperature attained during a 2-hour period. When the effects of 13 stimulating drugs (described earlier) were analyzed it was found for grouped animals that drug lethality was significantly and positively related to both motor activity and rectal temperature (Table 2). This was not true for isolated mice, again indicating that isolated and grouped animals respond differently to stimulating drugs. Importantly, motor activity and rectal temperature did not correlate significantly for grouped animals ($r = .48$), although the correlation was moderate and in the expected direction. This would suggest that the critical response is a rise in body temperature; although motor activity perhaps adds to the total effect by its action on body temperature (Hardinge & Peterson, 1963). Still unclear is how body temperature and neural arousal interact and to what degree these depend upon adrenal activation. As mentioned earlier, Swinyard and his colleagues (1961) found that amphetamine lowered the threshold for neural arousal but they were unable to find an associated change in body temperature. Further investigation may give considerable weight to functions of the central nervous system. Until such a time that the important features of group behavior can be outlined and appropriate weight assigned to separate physiological responses the correlational model suggested by Figure 1 can serve to point out important relations and critical interactions.

REFERENCES

ASKEW, B. M. Hyperpyrexia as a contributory factor in the toxicity of amphetamine to aggregated mice. *Brit. J. Pharmacol. Chemother.*, 1962, **19**, 245–257.

BARNETT, S. A. *A study in behaviour.* London, England: Methuen, 1963.

BERGER, F. M. Classification of psychoactive drugs according to their chemical structures and sites of action. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior.* New York: Wiley, 1960. Pp. 86–105.

BURN, J. H., & HOBBS, RONEEN. A test for tranquillizing drugs. *Arch. int. Pharmacodyn. Ther.*, 1958, **113**, 290–295.

CHANCE, M. R. A. Aggregation as a factor influencing the toxicity of sympathomimetic amines in mice. *J. Pharmacol. exp. Therapeut.*, 1946, **87**, 214–219.

CHANCE, M. R. A. Factors influencing the toxicity of sympathomimetic amines to solitary mice. *J. Pharmacol. exp. Therapeut.*, 1947, **89**, 289–296.

CHRISTIAN, J. J. Reserpine suppression of density-dependent adrenal hypertrophy and reproductive hypoendocrinism in populations of male mice. *Amer. J. Physiol.*, 1956, **187**, 353–356.

D'ARCY, P. F. The conditioning of experimental animals. *J. Pharm. Pharmacol.*, 1962, **14**, 411–415.

D'ARCY, P. F., & SPURLING, N. W. The effect of cortisol and corticotrophin on amphetamine toxicity in mice under crowded and non-crowded conditions. *J. Endocrinol.*, 1961, **22**, 25–36.

FINK, G. B., & LARSON, R. E. Some determinants of amphetamine toxicity in aggregated mice. *J. Pharmacol. exp. Therapeut.*, 1962, **137**, 361–364.

GREENBLATT, E. N., & OSTERBERG, A. C. Correlations of activating and lethal effects of excitatory drugs in grouped and isolated mice. *J. Pharmacol. exp. Therapeut.*, 1961, **131**, 115–119.

GROLLMAN, A. *Pharmacology and therapeutics.* Philadelphia, Pa.: Lea & Febiger, 1960.

GUNN, J. A., & GURD, M. R. The action of some amines related to adrenaline: Cychoherylalkylamines. *J. Physiol.*, 1940, **97**, 453–470.

HARDINGE, M., & PETERSON, D. I. The effect of exercise and limitations of movement on amphetamine toxicity. *J. Pharmacol. exp. Therapeut.*, 1963, **141**, 260–265.

HÖHN, R., & LASAGNA, L. Effects of aggregation and temperature on amphetamine toxicity in mice. *Psychopharmacologia,* 1960, **1**, 210–220.

HUNT, H. F., & BECKWITH, W. C. The effect of electroconvulsive shock under phenuron, dilantin, and amphetamine medication on the conditioned emotional response. *Amer. Psychologist,* 1956, **11**, 442. (Abstract)

LASAGNA, L., & MCCANN, W. P. Effect of "tranquilizing" drugs on amphetamine toxicity in aggregated mice. *Science,* 1957, **125**, 1241–1242.

McLean, J. R., & McCartney, M. Effect of *d*-amphetamine on rat brain noradrenaline and serotonin. *Proc. Soc. Exp. Biol. Med.*, 1961, **107**, 77–79.

Maxwell, D. R. The activity of some adrenaline antagonists in tests for tranquilizing action. In P. B. Bradley, P. Eniker, & C. Radouco-Thomas (Eds.), *Neuropharmacology.* Amsterdam, The Netherlands: Elsevier, 1959. Pp. 365–373.

Moore, K. E. Toxicity and catecholamine releasing actions of *d*- and *l*-amphetamine in isolated and aggregated mice. *J. Pharmacol. exp. Therapeut.*, 1963, **142**, 6–12.

Rothballer, A. B. The effect of catecholamines on the central nervous system. In O. Krayer (Ed.), *Catecholamines.* Baltimore, Md.: Williams & Wilkins, 1959. Pp. 494–547.

Sharp, J. C., Nielson, H. C., & Porter, P. B. The effect of amphetamine upon cats with lesions in the ventromedial hypothalamus. *J. comp. physiol. Psychol.*, 1962, **55**, 198–200.

Southwick, C. H., & Bland, V. P. Effect of population density on adrenal glands and reproductive organs of CFW mice. *Amer. J. Physiol.*, 1959, **197**, 111–114.

Swinyard, E. A., Clark, L. D., Miyahara, J. T., & Wolf, H. H. Studies on the mechanisms of amphetamine toxicity in aggregated mice. *J. Pharmacol. exp. Therapeut.*, 1961, **132**, 97–102.

Swinyard, E. A., Radhakrishnan, N., & Goodman, L. S. Effect of brief restraint on the convulsive threshold of mice. *J. Pharmacol. exp. Therapeut.*, 1962, **138**, 337–343.

Thiessen, D. D. Varying sensitivity of C57BL/Crgl mice to grouping. *Science*, 1963, **141**, 827–828.

Thiessen, D. D. Population density and behavior: A review of theoretical and physiological contributions. *Tex. Rep. Biol. Med.*, 1964, **22**, 266–314. (a)

Thiessen, D. D. Population density, mouse genotype, and endocrine function in behavior. *J. comp. physiol. Psychol.*, 1964, **57**, 412–416. (b)

Thiessen, D. D., & Rodgers, D. A. Population density and endocrine function. *Psychol. Bull.*, 1961, **58**, 441–451.

Thiessen, D. D., Zolman, J. F., & Rodgers, D. A. Relation between adrenal weight, brain cholinesterone activity, and hold-in-wall behavior of mice under different living conditions. *J. comp. physiol. Psychol.*, 1962, **55**, 186–190.

Turner, C. D. *General endocrinology.* Philadelphia, Pa.: Saunders, 1961.

Warren, M. R., & Werner, H. W. Acute toxicity of vasopressor amines: Part 1. Effect of temperature. *J. Pharmacol. exp. Therapeut.*, 1946, **86**, 280–283.

Weaver, L. C., & Kerley, T. L. Strain differences in response of mice to *d*-amphetamine. *J. Pharmacol. exp. Therapeut.*, 1962, **135**, 241–244.

Weiss, B., Laties, V. G., & Blanton, F. L. Amphetamine toxicity in rats and mice subjected to stress. *J. Pharmacol. exp. Therapeut.*, 1961, **132**, 366–371.

(Received February 19, 1964)

Psychological Bulletin
1964, Vol. 62, No. 6, 411–426

TAXONOMIES AND CORRELATES OF PHYSIQUE[1]

RICHARD G. DOMEY, JAMES E. DUCKWORTH, AND ANTHONY J. MORANDI

Harvard University School of Public Health

"Subjective" and "objective" taxonomies of human physique are more or less correlated among themselves and in different degrees with an array of nonanthropometric variables of interest to psychologists. There is probably closer correspondence between anthropometric data and function than is reported in the literature. The most advanced taxonomies can be improved, principally by: greater differentiation of predictive anatomical variables; increasing the scope of the systems to include such elements as internal body organs, the differential influences of nutrition, and age; and through use of refined multivariate statistical techniques for representing and analyzing complex data.

During the 1940s and early 1950s there was considerable interest evinced by psychologists in constitutional taxonomies because it was evident that systematic accounts of behavior depend, at least in part, upon systematic accounts of variation in physique. In the last few years this area of specialization seems to have attracted less attention. However, in other sciences, particularly physical anthropology, there has been virtually no diminution of either interest in or effort expended on this topic. It would be to the advantage of the psychologist to reconsider this theme once again, particularly the specialized areas of measurement, statistical representation, developing taxonomies, and the relationship of form to function. These are natural provinces for the behavioral scientists, for either the theoretician or for his applied psychology counterpart.

There are many unsolved methodological and technical problems associated with the development of taxonomies for representing physique. What measurement should be used; what are the best measures; how should certain lengths be defined? Can objective data and the prediction formulas be generalized for both sexes, all races, all ages? How do "objective" measures of physique correlate with "subjective" ratings of body type? What is the relationship of structure to performance? The questions are endless.

The work of Thurstone (1946a; 1946b, 1947) supports the contention that no statistical system, by itself, establishes the reality of discrete body types. This is a fundamental problem to be solved by appropriate research. On the other hand there is nothing to preclude the development of scaling systems for physique regardless of the underlying genetic variables and the response of the body to plenty and privation. Many scientists have exercised their prerogative to sort and classify differences among people throughout recorded history and will continue to do so. Nonetheless, the known taxonomies merely represent convenient ways of classifying individual constitutional differences according to more or less "standardized" schemata.

In practice, several different systems are commonly used, regardless of many reservations connected with typology. For example, differentiation among physiques, however dependent upon imper-

[1] This study was supported by Research Grant RD-592 from the Vocational Rehabilitation Administration, Washington, D. C.

fect taxonomic methodologies, has been studied and associated with age (Behnke, 1878; Clarke, Irving, & Heath, 1961; Hunt & Barton, 1959; Newman, 1952; Sheldon, Stevens, & Tucker, 1940), climate (Newman & Munro, 1955; Roberts, 1953), delinquency (Glueck & Glueck, 1950; Sheldon, Hartl, & McDermott, 1949), disease (Damon, 1960, 1962; Damon, Fowler, & Sheldon, 1955; Draper, Dunn, & Seegal, 1924; Pearl & Ciocco, 1934), endurance (Caldwell, 1964), maturation (Acheson & Dupertuis, 1957; Hunt & Barton, 1959; Karvonen & Kunnas, 1947; Krogman, 1956; Newman, 1952), motor ability and motor capacity (Bull, 1958; Caruth, 1953; Davies, 1957; Sills, 1950), nutrition (Jones, 1957; Keys, 1946; Lasker, 1947), occupation (Damon, 1955; Damon & McFarland, 1955; Garn & Gertler, 1950; Hannan, 1945; McFarland, Damon, & Stoudt, 1958), personality and behavior (Adcock, McCreary, Ritchie, & Somerset, 1958; Bull, 1958; DiGiovanni, 1919; Eysenck, 1953; Fiske, 1944; Heath & Seltzer, 1945; Parnell, 1958; Seltzer, 1946; Sheldon & Stevens, 1942; Wiersma, 1933), physical performance (Barter, Emanuel, & Truett, 1957; Bookwalter, 1952; Cullumbine, 1949; Duggar, 1963; Eränko & Karvonen, 1955; Parnell, 1951b; Pere, Kunnas, & Telkka, 1954; Sills & Mitchem, 1957; Sinelnikoff & Grigorowitsch, 1931; Tappen, 1950; Telkka, Pere, & Kunnas, 1951), psychomotor performance (Jones, 1957; Jones, 1949; Mohr & Gundlack, 1937; Sills & Everett, 1953), reaction time (Janoff, Geck, & Child, 1950; Smith & Bogarsky, 1943), smoking (Damon, 1961; Parnell, 1951a; Seltzer, 1959), specific gravity (Dupertuis, Pitts, Osserman, Welham, & Behnke, 1951), strength (Clarke, 1954, 1957; Clarke et al., 1961; Hunsicker, 1955; Hunsicker & Greey, 1957; Jones, 1947; Larson, 1940; Lookabaugh, 1937; Roberts, Provins, & Morton, 1959), susceptibility to psychological pathology of various types (Rees, 1950; Rees, 1961; Wertheimer & Hesketh, 1926), vital capacity (Allen, Peng, Cheng, Huang, Chang, & Tang, 1956), and weight (Behnke, 1961a; Behnke, Guttentag, & Brodsky, 1959; Bullen, 1963; Hechter, 1959; Stoudt, Damon, & McFarland, 1960; Tanner, 1951; Taylor & Behnke, 1961). Many other topical studies are not mentioned in this list.

Classification Systems

According to Pearl and Ciocco (1934) the study of body forms has, for one of its major objectives, the discovery of stable organic variables associated with it, and, eventually, the precise numerical measurement of such variables including their interrelationships. The main technical problem is not the invention of taxonomies, but rather the search for constants upon which taxonomies ultimately must depend. Still in the process of development, all classification systems are biased by many ephemeral dimensions. Confusing systems of classifying living objects with their genetic determinants lead only to useless controversy. One theme should not be mistaken for the other even though the available taxonomies are rather makeshift at best.

Physique, the body form of an individual, is more susceptible to measurements than almost any other constitutional or related factors such as psychological and physiological variables. The history of the classification of human constitution is summarized in detail by Ciocco (1936) and Sheldon, Stevens, and Tucker (1940). The earliest attempts to classify body build were essentially subjective evaluations. Hippocrates (undated) separated human

constitution into two general groups: the short, heavy, and strong; the tall, slight, and weak. He associated these physical categories with susceptibility to disease. Galenus (1881) related the four "humours" to body variations. Since Galenus' works remained the standard text for centuries, little was added to constitution analysis until early in the nineteenth century when Rostan (1828) proposed three body categories which seemed to be the prototypes of endomorphy, mesomorphy, and ectomorphy. Beneke (1878) suggested two distinct body groups associated with ectomorphy and endomorphy. He also measured the internal body organs. Variations in size were found to be associated with age, body classification, and disease. Subsequent work of the Italian anthropologists', especially Lombroso's (1889, 1911) studies of criminality and atavism, reinforced the correlations between behavior and physical constitution. Viola (1932) is also known for his work on body measurements.

Kretschmer (1921), a German psychiatrist, was responsible for a new wave of interest in the relationship between anthropology and psychology. Modifying the system posited by Rostan he classified physique into pyknic, asthenic, and athletic groups. Because of his dominant concern in the psychological aspects of constitution, he associated body form with psychological disorders. Accordingly, he differentiated between schizothymic and cyclothymic syndromes and attempted to relate these concepts to body types. Later, and unsuccessfully, he tried to correlate muscular and dysplastic body structure with epileptoid disorders. In contrast to this, he established certain associations between mental illnesses and physique which have not been disproved. The general result was that Kretschmer influenced many workers even though he has been criticized for his methods, subjectivity, and statistical analyses of his data.

Considered by many to be the most advanced, the next major schema was developed by Sheldon who modified the historical systems and attempted to construct a rationale around the facts of embryological development. For a comparison of Kretschmer's and Sheldon's systems see Ekman (1951). Sheldon retained Kretschmer's dysplasia variable and added gynandromorphy (g) and texture (t) indices. He (Sheldon et al., 1940; Sheldon, Dupertuis, & McDermott, 1954) asserted that the three germinal layers of the embryo, namely, the ectoderm, endoderm, and the mesoderm, give rise to three different kinds of tissue; significant differences in the proportions with which these tissues occurred in different physiques suggested different body types. This is a statistical inference for genetic determination of "type," nothing more, and thus cannot be considered an explanation of the taxonomy which provided the data upon which it depends. He rated each component on a 7-point scale of which, it is agreed, many permutations, such as 7-7-7, 1-1-1, or 6-1-6, could not logically exist. He also stated that any somatotype in which the sum of the three components is less than 9 or over 12 (revised to 13) should be reexamined by another observer.

A prime contribution to science was his standardized photoscopy technique. Also, in this same frame of reference is the relationship of photogrammetric anthropometry to an androgyny scale (Tanner, 1951) and Tanner's (1954) remarks on reliability of anthroposcopic somatotyping. Meredith (1940) and Humphreys (1957) have provided acute critiques of Sheldon's systems. Meredith questioned the sources of Sheldon's data,

the quantity, and the definitions of endomorphy and ectomorphy, while Humphreys contended that Sheldon's scale is more ipsative (individual) than normative (group). He also asserted that the arm chair choice of types "restricts the data in predictable ways [Humphreys, 1957, p. 227]."

Several other systems, mainly derived from Sheldon's work, have emerged. Hooton (1948, 1951, 1959) used the Sheldonian technique to rate endomorphic and mesomorphic components, and the ponderal index to rate "ectomorphy." Though his system is not in use today his report is a source book for much information on the constitution of soldiers. In his work are data on 35,000 soldiers from the original series of 50,000. In 1954 his death terminated the project, otherwise 9,000 to 10,000 more of the total series would have been evaluated. All these data are to be made available for further scientific study.

Since the correlation between these two classification schemes is known, the work retains its value even though the Hooton system is no longer used. A statistical comparison of these taxonomies was done by Dupertuis and Emanuel (1956), who found correlations in the .80s for each of the three major components, including the ectomorphy-ponderal index. However, the highest correlation (.86) was found for ectomorphy, the third component. Gynandromorphy correlated only .66 and dysplasia .05, a functional zero. On the average, the Hooton ratings were .51 unit higher than Sheldon's for endomorphy, the first component; 1.00 unit lower for mesomorphy, the second component; and .67 unit higher for ectomorphy, the third component. The findings on endo plus ecto agree quite well with Hooton's (1959) own statements:

> muscular development on the basis of this series is marked more strictly and more conservatively. The difference between our rating and that of Sheldon is probably nearly one grade. . . . In the appraisal of fatty development, on the other hand, it is certain that our grading is more liberal than that of Sheldon in the lower ranges of fatty development [p. 12a].

Among others, Parnell (1954a, 1954b) considered somatotyping to be an art, not a science. He also contended that somatotyping required too much labor, time, and expense. As an alternative, he used easily obtained physical measures, and by comparing these with the standard deviations above and below the respective means of the obtained values, he closely approximated somatotype ratings.

Later, Parnell (1958) refined his system by changing the original measures, including the favored ponderal index, and eliminating others which were not proportionately useful. Major changes were made in plotting the scales around their means with column units equivalent to .5 *SD*, a procedure which resulted in a 13-point scale which was extended to include the range of the female population. Parnell analyzed distributions of these measures on a simple basis of mean plus standard deviation and showed several dimensions of physique to be skewed. But if these data are truly skewed, then description of the distributions in terms of standard deviation is technically erroneous and misleading. An obvious improvement could have been achieved had skewed measures been transformed, normalized, or represented by more appropriate mathematics. The statistical treatment of these data was clearly limited and unnecessarily simplified.

Skewness is a real variable and to ignore this finding is to sacrifice a discovery which is worthy of the most careful study. Under certain circum-

stances skewness may not interfere in any practical way with establishing relative probabilities of similarities or difference among certain distributions. But when inherent in a descriptive taxonomy it must be considered a primary phenomenon. For example, skewness is clearly associated with body diameters, many of which are vulnerable to transitory influences. Measures of body and body extremities in units of length, however, appear to be "normally" distributed and far less vulnerable to impermanent external states. But why should one dimension be skewed and the other not?

Using logic similar to Parnell's, Lindegard (1953, 1956) developed a somewhat different system composed of four major factors: length, sturdiness, fat, and muscle. In this construct, length is represented by radial and tibial bone lengths and sturdiness by bone widths. Muscle is measured by dynamometers, hand grip, shoulder pull, and shoulder thrust. The measures are taken three times and in each case the highest is considered the best statistic. These three classes of scores are evaluated by means and standard deviations. A fat factor is determined by regression equations which include the amount of muscle and certain body circumferences.

Although Lindegard's method is among the more complicated, it has not resolved the problem of which dimensions give the most information. Since Lindegard has modified his system as certain measures have been found to yield more information than others, it must be concluded that the system is not yet standardized, a characteristic of all taxonomies, not Lindegard's alone.

A complex system devised by Schmitz and analyzed by Schich (1953) is essentially as follows. Transformation of height, chest, and abdominal circumference establishes three axes: I, II, III—corresponding to types, leptosome, athletosome, and pyknosome. Body weight is transformed to a cube 1,000 cubic centimeters with edge of 10 centimeters and weight of 1 kilogram. The formulae then become:

$$\mathrm{I} = 1 + w(h - w - 1)$$

$$\mathrm{II} = 1 - \frac{w}{2}(h - w - 1 - b + d)$$

$$\mathrm{III} = 1 - \frac{w}{2}(h - w - 1 + b - d)$$

where w = cube edge length, h = cube height, b = cube breadth, and d = cube depth expressed in logarithmic values of absolute measures. The product of the three axes always equals 1,000 cubic centimeters. Although it is impossible to equate Schmitz's evaluation with Kretschmer's (1921) types, the use of three-dimensional mathematics to represent physique is a scientific advance.

However, there are other issues. Clearly, temporary body states have plagued taxonomists and will continue to do so unless they are identified and integrated into classification systems. For example, Sheldon (Sheldon et al., 1940) asserted that body type is constant throughout life despite nutritional effects. But Howells (1957), in a study of a representative sample of Hooton's (1948) military data, pointed out that girth measures are very important in nutritive and age studies and that girth has been avoided if not actually rejected by physical anthropologists. Accordingly, Parnell's (1958) treatment of skewed data will illustrate this point. If nutritional or other change can disrupt the development of the body build to such an extent that the rating, made with the aid of the ponderal index, thereby varies significantly, then extant prediction data have limited application. Sheldon's data are in this category.

Certain starvation experiments, to which girth dimensions are vulnerable, are informative. Keys (1946), at the University of Minnesota, conducted a closely controlled 24-week famine diet study of 34 volunteers who were conscientious objects. Average weight loss was 24%. Seventeen body diameters suggested by Sheldon were measured. All body dimensions, including stature, markedly decreased. Endomorphy decreased an average of 49%, mesomorphy was reduced 43%, and ectomorphy increased 77%. Lasker (1947) also described the effects of partial starvation on somatotype and concluded that somatotype "serves better as a measure of nutritional status than as a measure of inherent tendencies to specific constitutional types [p. 324]."

For some, Lasker's work suggests that physical constitution itself is a highly variable entity, only appearing stable because of relative economic and sociological equilibrium of one culture or another. This condition suggests a phenotype, and not a morphogenotype, which is a structural genotype. But if this kind of proposition is accepted without qualification it implies that physique is so amorphous as to be immeasurable. For example, if this is true then whole areas of the science of physical anthropology must either be abandoned or a very different basis for classification of physique developed. Otherwise physical anthropometry and systems of sorting physique, basic tools in anthropology, are mere technical extravagances.

The effects of malnutrition do not sustain these types of rebuttals to the various body classification systems. The starved mesomorph is not comparable to the well-nourished mesomorph, and no such comparison should be made either to support or reject the body-type hypothesis. Instead, more attention should be given to discriminating among constant and inconstant variables. The failure to do so is at the root of many technical difficulties associated with anatomical classification systems of all types. Starvation experiments are irrelevant except as they demonstrate the organism's response to feeding, which, it is suggested, is predictable. More to the point would have been the development of a mathematical scheme expressing the nature of change in dimensions of physique as functions of feeding and deprivation. The complexity of taxonomies would increase, but that is all.

The sort of argument proposed by Lasker and others is largely self-defeating precisely because, in turn, they do not propose any definition of "normality" and because they end by demonstrating the possibility of classifying malnourished men using one of Sheldon's several systems. In a sense, reliability of Sheldon's system was demonstrated since changes in body conformation during starvation were reflected by modification of somatotype ratings. On the other hand, modification of somatotype occurred only because no differentiation was made between the starved and nonstarved subjects.

Without further entering the controversy over age and nutritional effect on somatotype, that is, whether the somatotype is a phenotype or morphogenotype, it is suggested that these questions (and many others) might be resolved through the study of large-scale samples made possible by an easier, standardized method of somatotyping.

The utility of these concepts, Brozek (1963) suggested, may be increased by a change in direction of research, not solely by improvement in technique. He asserts that the somatotype is not sufficiently refined as a research tool, and that the next major advance in constitutional studies will be the analysis of internal body form for which

Beneke (1878) had set precedence. Damon (1962) provided a summary of some of the work relating somatotype to various physiological functions and pathological disturbances, two other avenues of study.

Obviously, in somatotyping one of the major problems in the past has been the diversity of classification systems used in the investigations of a specific problem area. The result of this seems to be that no one system has shown a clear-cut advantage over another. Thus, overall interpretation of obtained data is obscured by methodological and technical differences, as well as varying criteria. Even somatotype evaluation systems are in a state of flux, so reliability of data obtained in different laboratories is difficult to establish. Those who classify through statistical systems are completely at the mercy of the original measures which are seldom standardized from one study to another. The results from several competitive schemes, nevertheless, are remarkably similar, a finding which has encouraging, if obvious, implications for research.

Anthropometric Correlates of Body Classifications

Other researchers have attempted to construct more comprehensive systems which would link the somatotype directly to more extensive arrays of anthropometric measures (Damon, Bleibtreu, Elliot, & Giles, 1962; Dupertuis, 1950; Hjortsgo & Lindegard, 1953; Parnell, 1954a; Wigant, 1933). Once the relationships are established, the obvious advantages are the reduction in time, effort, equipment, and specialized training which are necessary parts of the present somatotyping methodology since the more objective body measures could be substituted for subjective ratings.

Several methods for representing such a relationship have been suggested. Each system delineates a schema for classifying body build in gross categories by simple and quick measures. The first was the body index and its variation (Pearl, 1940); then factor analyses (Burt, 1944, 1947, 1949; Hammond, 1942, 1957a, 1957b; Hempel & Fleishman, 1955; Howells, 1951, 1952; Spearman, 1927; Thurstone, 1946a, 1946b, 1947); and finally, multiple-regression equations (Caruth, 1953; Damon et al., 1962; Jones, 1947, 1949). Obviously, the ultimate interpretations of results obtained by these and other statistical methods depend upon the mathematical assumptions which subsume the types of analyses used. For example, skewed anthropometric data are generally treated as if skewness was unimportant. Also body conformation may not, and probably does not, fit the factorial-analysis restriction of orthogonality. Furthermore, the data per se selected for analysis are also codeterminants. And accumulating experience appears to indicate that any really acceptable taxonomy is likely to be complex.

Although many are no longer in use, a list of some 30 body indices, and the way in which each is calculated, is presented by Tucker and Lessa (1940). Most of these depend upon one or more measures of height and breadth, or weight, mathematically combined to yield a number. These obtained composite values may be considered indices occupying relative positions in one or another type of statistical distribution. The relationship of indices to the evaluation of physique has been explored by Rees (1949).

The Rees-Eysenck body index (Eysenck, 1959; Rees & Eysenck, 1945) is used commonly as a classification referent. Their contention is that factorial analysis of physique suggests that the body be regarded as a rectangle which can be described with fair accuracy by two independent dimensions,

height and width. By multiplying the two, total size is shown; dividing one by the other yields a ratio representing shape. Using height and transverse chest width, and evaluating obtained means and standard deviations, the two indices used in conjunction describe both general size and shape of the body.

There is a definite, clear-cut, "high" correlation between this obtained number system and subjective classifications of physique: Ectomorphy has high indices of body build; endomesomorphy low indices. However, the data at the lower end of the index scale could be grouped in two classes. For example, a person could be stocky for different reasons, muscularity or fatness, or perhaps a combination of both. This relationship is not revealed by any known index, including the Rees-Eysenck index. These authors, and others, contend however, that the fat-muscle ratio of physique is not a stable factor. Damon and Goldman (1964), on the other hand, evaluated 10 densitometric equations inclusive of anthropometric measures and found the best predictors to be triceps and subscapula skinfolds. Brozek (1963) has emphasized the importance of more detailed study of internal body structure and function as these relate to variations among constitutional types.

Among the commonly used ratios is the ponderal index, $\frac{\text{height}}{\sqrt[3]{\text{weight}}}$, a formulation that gained more acceptance than others. Height, of course, varies little after maturity, and the use of the $\sqrt[3]{\text{weight}}$ as the denominator assures that small deviations in body state will be represented by a change of rated position in the population distribution.

The ponderal index is also a central part of the Sheldonian rationale. In this schema a 5-year interval age table has been constructed. By calculation of the ponderal index used as a generalized referent, a given body type may be "identified." There is usually a range of possible somatotypes as a function of age for a given ponderal index, except at the extremes of the distribution. Careful analysis of Sheldon's (Sheldon et al., 1940) tables, however, reveals a general lack of differentiation of somatotype by the ponderal index, which, at best, serves only as a gross guideline for subjective ratings. If a body type, as rated by the viewer, does not agree with the body classification table, the rating must be recalculated. Since the somatotype must be "rescored" and made consistent with the ponderal index, the established relationship is circular. With license to change interpretation, post hoc, reliability of prediction may be expected to closely approximate unity, and it does. In spite of these, and other technical embarrassments, no valid argument can be raised against the central idea of a taxonomy for physique at any level in the phylogenetic scale, nor against the utility of such taxonomies as convenient, though arbitrary, generalizations.

One problem which has limited the use of indices also has limited the use of objective methods of classifying physique as well, that is, identifying the most appropriate constitutional variables. For this purpose, factor analysis was first used as an analytic method in physical anthropology by Spearman (1927). But in his factor-analysis programs no suitable procedures were included for determining the amount of information independently contributed by any given measure or ratio of measures. Factor analysis as refined by Thurstone (1940) utilized statistical theory and techniques not available to Spearman. Thurstone (1946b) evaluated data originally obtained by Hammond (1942) and applied factor-analysis methods to the intercorrelations of 12 anthropometric measurements. By rotation of the axes these were shown to yield a simple structure of four

factors: head size, trunk size, or girth, and extremity size. The length and girth factors were also correlated with each other (.38), but the other two variables seemed to be relatively independent. Thurstone perceived the distribution of body types as a continuum and ended his work by disavowing belief in discrete types. Of course, anatomical dimensions are not necessarily distributed according to a number of assumptions underlying probability statistics in general. More recent improvements in analytic methods are likely to advance the development of statistical physical anthropology.

When somatotyping was done by Sheldon's method, and by direct measurement as well, on 154 Oxford men, 16–20 years of age, Parnell (1954a) found that 90% of the ratings corresponded within .5 unit or better. For a further study, 292 male Oxford undergraduates were used to validate the deviation tables against photoscopic ratings. It was found that 87.3% of the ratings were correct within .5 unit, although there was a tendency to rate mesomorphy lower by about .2 unit with the photoscopy method. Lindegard's system (1953, 1956) should also be considered in this context as already mentioned.

There are specialized studies of distribution extremes such as those by Dupertuis (1950), Howells (1951, 1952), and Liu (1952) about which a word of caution is offered. The method and practice of comparing group extremes have statistical disadvantages and should be avoided. For example, the proper statistical universe is the sample data which may, as distributions, be shown to have very different properties from extreme measures. This means that generalization from exotic, nonrandom samples-of-samples to the larger population is virtually impossible. This practice imposes unnecessary limitations which abrogate the apparent convenience of selecting extremes for experimental study.

Dupertuis (1950) selected three groups of 10 members each in the three extreme somatotypes from the general New York City, male, 21–25 year old population. He attempted to distinguish among the subjective ratings of the three groups by physical measures. Using the means of his measurements he found that ectomorphs were easily differentiated from the other two groups. Endomorphs and mesomorphs differed mainly in trunk diameters and circumferences. Head, face, body, and limb lengths failed to delineate these groups to a significant degree. Dupertuis then concluded that extreme anthropometric measurements were only "moderately" successful in identifying differences among extreme somatotypes and thus would be of questionable value in differentiating the midrange and closely related somatotypes. Although there are technical problems associated with comparative samples composed of extremes, this outcome did not contradict Thurstone's conclusion suggested by factor analysis of anthropometric data.

Howells selected five men from each of the extremes of Dupertuis' groups and subjected their data to factor analyses. He was able to identify three factors which did not correspond to Sheldon's three major components: padding between the skin and bones, top heaviness versus bottom heaviness, and trunk circumference development versus limb development. He concluded that mesomorphs and endomorphs were indistinguishable through anthropometric measurement alone. However, these data are probably not appropriate to this generalization since differences could exist in the remainder of the distributions not studied.

Two prime questions are still whether the evaluation of the somatotype has been exhausted and whether any fur-

ther refinements in somatotyping will increase the correlations with psychological, physiological, and pathological states. Even without resolving these questions the presumed somatotype usefulness can be increased by merely reducing the time, effort, and equipment necessary to obtain reliable estimates of somatotype. This, of course, was one reason for devising body indices: Parnell's system, the use of factor analysis, and application of other types of statistical analysis of which the work of Damon et al. (1952, 1958, 1962) represents a recent attempt. These methods were used to introduce greater objectivity into somatotyping as well as the simplification of the total procedure.

A more complete and systematic differentiation among anthropometric representations of fat and muscle would, of course, increase the predictability of mesomorphy and endomorphy as they are subjectively rated. These two measures are the least accurately predicted of the three major somatotype categories. Clearly, not all possible analytic statistical avenues have been utilized, especially the analysis of body constellations and the degree that these may predict subjective ratings. It is suggested that further advances can be made, not only by the examination of internal body form, and physiological phenomena and function, but by a reexamination of statistical effects through more advanced procedures.

Taxonomists, it seems, may be classified in a general way as subjective or objective systematizers. Historical processes and propensities for preferring one approach to the other have led to unnecessary conflicts which, in effect, could be resolved by simple compromises. Those who prefer direct measurement of physical variables tend to claim precedence in the name of objectivity since it is difficult to deny the tape measure, anthropometer, and scales. Those who classify by standardized perceptual techniques claim precedence by asserting that the eye can discriminate among variables not easily recorded by available mechanisms for direct measurement. But either group, in effect, supports the principles put forward by the other since the "subjectivists" resort to direct measurement in order to defend their taxonomies and the "objectivists" compare their constructs with the subjectivist's systems, thus, each one "validating" his concepts by resorting to the other. The psychologist is not surprised to observe that there is a far greater-than-chance correlation between the taxonomies of the two groups. But both groups must consider that the most effective system will probably be devised by those who find some way of integrating the best skills of each.

Physique, Biomechanical Performance, and Occupation

Clearly, the next important concept, once the possibility of classifying physique has been established, is whether physical structure and function are related. Common experience demonstrates a close relationship among the more obvious characteristics of structure and behavior. But general knowledge of this interaction could be greatly extended. This topic may be studied in several ways among which are observations of relevant physiological variables, the dynamics of biomechanics, and the manifestation of performance in the more diffuse behavioral endeavors of recreation and occupation. A topical bibliography has been assembled by Hansen, Cornog, and Hertzberg (1958). Nevertheless, the literature on the relationship of physique to performance and occupation is not extensive nor experimentally adequate. More reliable and valid studies will be appreciated and used.

There is, Edwards (1960) stated, a physiochemically determined upper-limit asymptote for muscle strength per unit of cross-section area. He predicted a curve for strength which he said should be followed more closely by persons who were nearest to their maximum potential development. Highly trained weight lifters performed as predicted "quite well," novice lifters "slightly," while the general population failed to approximate the predicted curve. Performance scores obtained from novice weight lifters, regional champions, and various world record holders were also compared. Champions and record holders reproduced the expected data except for the heavyweights whose fat was not found to be a detriment to performance. In measurements of fine movements, A. Jones (1957) found that obese and nonobese were not differentiated on all tasks.

Eleven extreme endomorphs, 11 extreme mesomorphs, and 7 extreme ectomorphs were examined by Liu (1952). He measured the muscle cross section in the upper arm and the dynamometric strength of the arm. The mesomorphs were stronger, absolutely and relatively. Ectomorphs and endomorphs were almost the same, provided the results were corrected for over- and underweight. Without this correction, the obtained arm values were: mesomorphs, 8.487 kg/cm^2; ectomorphs, 8.012 kg/cm^2; and endomorphs, 7.513 kg/cm^2. Remarks about the use of statistical extremes have already been made (also see Roberts, 1953; Roberts et al., 1959).

Using 62 Springfield College male students, 20–26 years old, Clarke (1954) found that the girth of the flexed-tensed arm was highly related to arm strength criteria. He (Clarke, 1957) also tried to relate anthropometric measures to physical performances involving trunk and legs. Clarke et al. (1961) associated somatotype with skeletal maturity, structural characteristics, and muscular strength. Their subjects were 259 boys, 37 for each year, aged 9–15 inclusively. Both types were classified into five groups according to predominance of endomorphy, mesomorphy, endomesomorphy, and balanced or midrange types. Endomorphs and endomesomorphs were largest in several torso and extremity girth measures. Mesomorphs were highest in both gross and relative strength and in muscular endurance. Ectomorphs and midrange types followed somewhat the same pattern. Sitting height, leg length, classification index, and lung capacity showed no differences among the groups. For their chronological age a significantly greater number of endomesomorphs were physically advanced than were retarded. The reverse was true for the midrange types.

Everett and Sills (1952) obtained scores and ratings from 400 subjects and related strength to stature, somatotype components, and anthropometric measurements. These subjects, ranging in age from 14 through 29, of whom 94% were under 20 years of age, were all classified by the Sheldonian somatotyping method. Weight, height, hand width, and mesomorphy were the most influential variables in the prediction of hand-grip strength (also see Sills & Mitchem, 1957; Tanner, 1952).

A longitudinal study of 89 boys and 87 girls was conducted by Jones (1947, 1949) who recorded growth data semiannually from the time each person was 11 years old until each reached 17 years and 6 months. Strength was correlated highest with weight, next highest with a combination of height and mesomorphy, not at all with endomorphy, and inversely with ectomorphy. The multiple correlation of these five factors contrasted with experiential or training factors in determining individual differences in strength.

Using 18 male and 18 female college students, Caldwell (1964) discovered that endurance was not significantly different among subjects when individual differences in strength were statistically held constant. However, when relative load was increased from 25% to 100% of maximum strength, rate of response endurance decreased from 252 seconds to 2 seconds. His six anthropometric referents were height, weight, upper and forearm lengths, and upper and forearm girths. Unlike other workers (Clarke et al., 1961), Caldwell distinguished between strength and endurance.

Within the 18–50 age range Lookabaugh (1937) predicted the total potential strength of adult males from their skeletal build; Roberts et al. (1959) showed a clear relationship between arm strength and body size, and among limb dimensions, especially arm girths.

In a review of occupational anthropology including 99 references, Damon and McFarland (1955) conclude with Simonson (1947) that exhaustive studies of the correlates of constitution are pertinent to industrial research. Although their analyses showed some relationship between occupation and physique, this was not consistently associated with physical expenditure. However, in other research involving 100 workers in an eastern factory, Garn and Gertler (1950) found that when workers were compared with a subgroup of 20 research workers from the same group, the researchers were more actomorphic than the total group by 45% to 23%. They conclude that selection of occupation is partially related to strength requirements and temperamental correlates of body type.

Bus and truck drivers tend to be endomorphic-mesomorphic and degree of success is associated with degree of mesomorphy. In the words of Damon and McFarland (1955) "heavy vehicle driving is a job of, if not for, the mesomorph low in gynandromorphy, and that the best drivers are even more mesomorphic [p. 734]." Similarly, aviators tend to be predominately mesomorphic, and again, success is related to degree of mesomorphy (Damon, 1955; McCormack, 1947; McFarland, Graybiel, Liljencrantz, & Tuttle, 1939).

The accumulated evidence suggests there is a greater-than-chance correspondence among body classification systems and that, in turn, these systems can be statistically associated with a wide variety of events. However, the degrees of relationship established are highly variable and are sufficient for only limited decision making.

Remarks

Examination of the literature shows, in general, that the statistical methods and techniques commonly used for representing the several taxonomies and for analysis of the anthropometric data lag behind the "state of the art." For example, most studies do not employ multivariate-analysis techniques, such as discriminate analysis, canonical analysis, stepwise, inductive regression analysis, and others. The greater analytic power of these techniques allows for more definitive handling of complex data than has been possible heretofore. Furthermore, the greater sensitivity of these techniques generally results in isolating effects not hitherto noticed, and in the cases of the more complex correlation techniques, revealing relationships of higher correspondence among related variables than previously have been obtained. The assumption is that anthropometric data, the most stable foundation for taxonomies of physique, are probably more highly associated among themselves and with other variables than has been reported.

This does not imply that the avail-

able taxonomies are sufficient since it is clear that many interacting variables such as muscle and fat simultaneously incorporated in body circumference measures fail to differentiate among physiques that are indeed very different. The failure to account for the influences of nutrition upon body conformation is another embarrassment to system building. But the point made here is that taxonomies dependent upon physical anthropometric data are by no means at the end of their development. Obviously, as these schema incorporate more variables, such as measures of internal body organs, physiological factors, and biomechanics, the predictive capacity of the systems will increase. So the suggestion is that modern statistical techniques coupled with more differentiated taxonomies will result in an increase of statistical concordance among many variables, and hence in general usefulness in many areas of decision making.

REFERENCES

ACHESON, R. M., & DUPERTUIS, C. W. The relationship between physique and rate of skeletal maturation in boys. *Hum. Biol.*, 1957, **29**, 167–193.

ADCOCK, C. J., MCCREARY, J. R., RITCHIE, J. E., & SOMERSET, H. C. A. Personality and physique. *Victoria U. Wellington. Publ. Psychol.*, 1958, No. 12.

ALLEN, T. H., PENG, M. T., CHEN, K. P., HUANG, T. F., CUANG, C., & TANG, H. S. Similarity of vital capacity in terms of body weight less adiposity in both sexes. *Metabolism*, 1956, **5**, 353–358.

BARTER, J. T., EMANUEL, I., & TRUETT, B. A statistical evaluation of joint range data. *USAF WADC tech. Note*, 1957, No. 57-311.

BEHNKE, A. R. Anthropometric tractination of body weight. *J. appl. Physiol.*, 1961, **16**, 949–954. (a)

BEHNKE, A. R. Quantitative assessment of body build. *Amer. J. Physiol.*, 1961, **201**, 960–968. (b)

BEHNKE, A. R., GUTTENTAG, O. E., & BRODSKY, C. Quantification of body weight and configuration from anthropometric measurements. *Hum. Biol.*, 1959, **31**, 213–234.

BENEKE, F. W. *Die anatomische Grundlagen der Konstitutionsanomalien des Menschen.* Marburg, 1878.

BOOKWALTER, K. W. The relationship of body size and shape to physical performance. *Res. Quart.*, 1952, **23**, 271–279.

BROZEK, J. Quantitative description of body composition: Physical anthropology's "fourth" dimension. *Curr. Anthropol.*, 1963, **4**, 3–39.

BULL, K. R. An investigation into the relationship between physique, motor capacity and certain temperamental traits. *Brit. J. educ. Psychol.*, 1958, **28**, 149–154.

BULLEN, B. Physical inactivity as a factor associated with obesity in adolescent girls. Unpublished doctoral dissertation, Harvard University School of Public Health, 1963.

BURT, C. The factorial study of physical types. *Man*, 1944, **54**, 82–86.

BURT, C. Factor analyses and physical types. *Psychometrika*, 1947, **12**, 171–188.

BURT, C. Subdivided factors. *Brit. J. Psychol. statist. Sect.*, 1949, **2**, 41–63.

CALDWELL, L. S. Relative muscle loading and endurance. *J. engng. Psychol.*, 1964, **2**, 155–161.

CARUTH, W. A. An analysis of motor ability and its relationship to constitutional body patterns. Unpublished doctoral dissertation, New York University, 1953. (*Dissert. Abstr.*, **13**:350)

CIOCCO, A. The historical background of the modern study of constitution. *Bull. Inst. Hist. Med. Johns Hopkins U.*, 1936, **4**, 23–28.

CLARKE, H. H. Relationship of strength and anthropometric measures to various arm strength criteria. *Res. Quart.*, 1954, **25**, 134–143.

CLARKE, H. H. Relationship of strength and anthropometric measures to physical performances involving the trunk and legs. *Res. Quart.*, 1957, **28**, 223–232.

CLARKE, H. H., IRVING, R. N., & HEATH, B. H. Relation of maturity, structural and strength measures to the somatotypes of boys through 15 years of age. *Res. Quart.*, 1961, **32**, 449–460.

CULLUMBINE, H. Relationship between body build and capacity for exercise. *J. appl. Psychol.*, 1949, **2**, 155–168.

DAMON, A. Physique and success on military flying. *Amer. J. phys. Anthropol.*, 1955, **3**, 217–252.

DAMON, A. Host factors in cancer of the breast and uterine cervix and corpus. *J. Nat. Cancer Inst.*, 1960, **24**, 483–516.

DAMON, A. Constitution and smoking. *Science*, 1961, **134**, 339–341.

DAMON, A. Some host factors in disease: Sex, race, ethnic group, and body form. *J. Nat. Med. Ass.*, 1962, **54**, 424–431.

DAMON, A., BLEIBTREU, H. K., ELLIOT, O., & GILES, E. Predicting somatotype from body measurements. *Amer. J. phys. Anthropol.*, 1962, **20**, 461–473.

DAMON, A., FOWLER, E. P., & SHELDON, W. H. Constitutional factors in otosclerosis and Menieres disease. *Trans. Amer. Acad. Ophthalmol. Otolaryngol.*, 1955, **29**, 444–458.

DAMON, A., & GOLDMAN, R. P. Predicting fat from body measurements: Densitometric validation of ten anthropometric equations. *Hum. Biol.*, 1964, **36**.

DAMON, A., & MCFARLAND, R. A. The physique of bus and truck drivers: With a review of occupational anthropology. *Amer. J. phys. Anthropol.*, 1955, **13**, 711–742.

DAVIES, E. A. Relationship between selected postural divergencies and motor ability. *Res. Quart.*, 1957, **28**, 1–3.

DIGIOVANNI, A. *Clinical commentaries deduced from the morphology of the human body*. New York, 1919.

DRAPER, G., DUNN, H. L., & SEEGAL, D. Studies in human constitution. *J. Amer. Med. Ass.*, 1924, **82**, 431–434.

DUGGAR, B. C. Continuous control performance of the human upper extremity. Unpublished doctoral dissertation, Harvard University School of Public Health, 1963.

DUPERTUIS, C. W. Anthropometry of extreme somatotypes. *Amer. J. phys. Anthropol.*, 1950, **3**, 67–85.

DUPERTUIS, C. W., & EMANUEL, I. A statistical comparison of the body typing methods of Hooten and Sheldon. *USAF WADC tech. Rep.*, 1956, No. 56-366.

DUPERTUIS, C. W., PITTS, G. C., OSSERMAN, E. F., WELHAM, W. C., & BEHNKE, A. R. Relation of specific gravity to body build on a group of healthy men. *J. appl. Physiol.*, 1951, **3**, 676–680.

EDWARDS, W. E. Relationship of human size to strength. *Amer. J. phys. Anthropol.*, 1960, **18**, 358. (Abstract)

EKMAN, G. *On the number and definitions of dimensions in Kretschmer's and Sheldon's constitutional systems: Essays in psychology*. Stockholm, Sweden: Almquist, 1951.

ERÄNKO, O., & KARVONEN, M. J. Body type of Finnish champion lumberjacks. *Amer. J. phys. Anthropol.*, 1955, **13**, 331–341.

EVERETT, P. W., & SILLS, F. Relationship of grip strength to stature, somatotype components and anthropometric measurements of the hand. *Res. Quart.*, 1952, **23**, 161–166.

EYSENCK, H. J. *The structure of human personality*. London, England: Methuen, 1953.

EYSENCK, H. J. The Rees-Eysenck body index and Sheldon's somatotype system. *J. ment. Sci.*, 1959, **105**, 1053–1058.

FISKE, D. W. A study of relationship to somatotype. *J. appl. Psychol.*, 1944, **28**, 504.

GALENUS, DET. *De temperamentis*. Cambridge, England: Macmillan & Bowers, 1881.

GARN, S. M., & GERTLER, M. M. An association between type of work and physique in an industrial group. *Amer. J. phys. Anthropol.*, 1950, **8**, 387–397.

GLUECK, S., & GLUECK, H. *Unraveling juvenile delinquency*. New York: Commonwealth Fund, 1950.

HAMMOND, W. H. An application of Burt's multiple general factor analyses to the delineation of physical types. *Man*, 1942, **52**, 4–11.

HAMMOND, W. H. The constancy of physical types as determined by factorial analysis. *Hum. Biol.*, 1957, **29**, 40–61. (a)

HAMMOND, W. H. The status of physical types. *Hum. Biol.*, 1957, **29**, 223–241. (b)

HANMAN, B. Matching the physical characteristics of workers and jobs. *Industr. Med.*, 1945, **14**, 405–430.

HANSEN, R., CORNOG, D. Y., & HERTZBERG, H. T. E. Annotated bibliography of applied physical anthropology in human engineering. *USAF WADC tech. Rep.*, 1958, No. 56.

HEATH, C. W., & SELTZER, C. C. *What people are*. Cambridge: Harvard Univer. Press, 1945.

HECHTER, H. Relationship between weight and some anthropometric measurements in adult males. *Hum. Biol.*, 1959, **31**, 235–243.

HEMPEL, W. E., JR., & FLEISHMAN, E. A. A factor analysis of physical proficiency and manipulative skill. *J. appl. Psychol.*, 1955, **39**, 12–16.

HIPPOCRATES. *On ancient medicine: The genuine works of Hippocrates*. (Trans. by F. Adams) New York: Wood, undated.

HJORTSJO, C. H., & LINDEGARD, B. Critical aspects of the use of indices in physical anthropology. *K. fysiogr. Sällsk. Lund Förh.*, 1953, **23**(8).

HOOTON, E. A. Body build in relation to military function in a sample of the United States Army. Cambridge: Harvard University, Department of Anthropology, 1948.

HOOTON, E. A. *A handbook of body types in the United States Army*. Cambridge: Harvard University, Department of Anthropology, 1951.

HOOTON, E. A. Body build in a sample of the United States Army. *USA Qm. Res. Engng. Cent. Environ. Protect. Res. Div. tech. Rep.*, 1959, No. EP-102.

HOWELLS, W. W. Factors of human physique. *Amer. J. phys. Anthropol.*, 1951, **9**, 159–191.

HOWELLS, W. W. A factorial study of constitutional type. *Amer. J. phys. Anthropol.*, 1952, **10**, 91–118.

HOWELLS, W. W. *Variation of external body form in the individual.* Cambridge: Harvard University, Peabody Museum, 1957.

HUMPHREYS, L. G. Characteristics of type concepts with special reference to Sheldon's typology. *Psychol. Bull.*, 1957, **54**, 218–228.

HUNSICKER, P., & GREEY, R. Studies in human strength. *Res. Quart.*, 1957, **28**, 109–122.

HUNSICKER, P. A. Army strength at selected degrees of elbow flexion. *USAF WADC tech. Rep.*, 1955, No. 54-548.

HUNT, E. D., JR., & BARTON, W. H. The inconstancy of physique in adolescent boys and other limitations of somatotyping. *Amer. J. phys. Anthropol.*, 1959, **17**, 27–35.

JANOFF, I., BECK, L. H., & CHILD, I. L. Relation of somatotype to reaction time, resistance to pain and expressive movement. *J. Pers.*, 1950, **18**, 454–460.

JONES, A. A study of the relationship between psychomotor performance and the physiological condition of obesity. *Dissert. Abstr.*, 1957, **17**, 1134–1135.

JONES, H. E. The relationship of strength to physique. *Amer. J. phys. Anthropol.*, 1947, **5**, 29–34.

JONES, H. E. Motor performance and growth: A developmental study of static and dynamometric strength. *Child Develpm.*, 1949, **1**, 182.

KARVONEN, A., & KUNNAS, M. On the age and constitutional type of the best Finnish lumbermen. *Ann. Acad. Sci. Fenn. Ser. A*, 1947, **16**, 1–2.

KEYS, A. Human starvation and its consequences. *J. Amer. Diet. Ass.*, 1946, **22**, 582–587.

KRETSCHMER, E. *Körperbau und Charakter.* Berlin, Germany: Springer, 1921.

KROGMAN, W. M. The physical growth of children: An appraisal of studies 1950–55. *Monogr. Soc. Res. Child Develpm.*, 1956, **20**, 1–91.

LARSON, L. A. A factor and validity analysis of strength variables and tests with a test combination of chinning, dipping and vertical jump. *Res. Quart.*, 1940, **11**, 82–96.

LASKER, G. W. The effects of partial starvation on somatotype. *Amer. J. phys. Anthropol.*, 1947, **5**, 323–342.

LINDEGARD, B. Variations in human body build. *Acta psychiat. neurol. Scand., Copenhagen*, 1953, Suppl. No. 96.

LINDEGARD, B. (Ed.) Body build, body function and personality. *Lunds U. Årsskr. Avd. 2*, 1956, **52**(4).

LIU, TIEN-SI. Muscular strength in relation to body type. Unpublished doctoral dissertation, University of Iowa, 1952.

LOMBROSO, C. *L'uomo deliquente.* Torina, 1889.

LOMBROSO, C. *Crime, its causes and remedies.* Boston, 1911.

LOOKABAUGH, G. The prediction of total potential strength of adult males from skeletal build. *Res. Quart.*, 1937, **8**, 103–108.

MCCORMICK, M. Y. *Physical characteristics of the 1946 airline transport pilot population.* (Medical Service Publication) Washington, D. C.: United States Department of Commerce, Civil Aeronautics Administration, 1947.

MCFARLAND, R. A., DAMON, A., & STOUDT, H. Anthropometry in the design of driver workspace. *Amer. J. phys. Anthropol.*, 1958, **16**, 1–24.

MCFARLAND, R. A., GRAYBIEL, A., LILJENCRANTZ, E., & TUTTLE, A. D. An analysis of the physiological and psychological characteristics of 200 civil airline pilots. *J. aviat. Med.*, 1939, **10**, 2–52.

MEREDITH, H. V. Comments on the varieties of human physique. *Child Developm.*, 1940, **2**, 301–309.

MOHR, G. J., & GUNDLACK, R. H. The relation between physique and performance. *J. exp. Psychol.*, 1927, **10**, 117–157.

NEWMAN, R. W. Age changes in body build. *Amer. J. phys. Anthropol.*, 1952, **10**, 75–90.

NEWMAN, R. W., & MUNRO, E. H. The relation of climate and body size in U. S. males. *Amer. J. phys. Anthropol.*, 1955, **13**, 1–18.

PARNELL, R. W. Smoking and cancer. *Lancet*, 1951, **1**, 963. (a)

PARNELL, R. W. Some notes on physique and athletic training. *Brit. med. J.*, 1951, **1**, 1292–1295. (b)

PARNELL, R. W. Physique and performance: Honours class at Oxford. *Brit. med. J.*, 1954, **2**, 49–66. (a)

PARNELL, R. W. Somatotyping by physical anthropometry. *Amer. J. phys. Anthropol.*, 1954, **12**, 209–240. (b)

PARNELL, R. W. *Behavior and physique: An introduction to practical and applied somatometry.* London, England: Arnold, 1958.

PEARL, R. An index of body build. *Amer. J. phys. Anthropol.*, 1940, **26**, 315–348.

PEARL, R., & CIOCCO, A. Studies on constitution: II. Somatological differences associated with diseases of the heart in white males. *Hum. Biol.*, 1934, **6**, 650–713.

PERE, S., KUNNAS, M., & TELKKA, A. Correlation between performance and physique in

Finnish athletes. *Amer. J. phys. Anthropol.*, 1954, 12, 201–208.

REES, L. The value of anthropometric indices in the assessment of body build. *J. ment. Sci.*, 1949, **95**, 171–179.

REES, L. Body size, personality and neurosis. *J. ment. Sci.*, 1950, **96**, 168–180.

REES, W. L. Constitutional factors and abnormal behaviour. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology*. New York: Basic Books, 1961. Ch. 9.

REES, W. L., & EYSENCK, H. J. A factorial study of some morphological and psychological aspects of human constitution. *J. ment. Sci.*, 1945, **41**, 8–21.

ROBERTS, D. F. Body weight, race, and climate. *Amer. J. phys. Anthropol.*, 1953, **11**, 533–558.

ROBERTS, D. F., PROVINS, K. A., & MORTON, R. J. Army strength and body dimensions. *Hum. Biol.*, 1959, **31**, 334–343.

ROSTAN, L. *Cours elementaire d'hygiene*. Paris: 1828. 2 vols.

SCHICH, C. Neure Methode der Körperbauanalyse nach Karl Ludwig Schmitz. [New method of body build analysis according to Karl Ludwig Schmitz.] *Z. KonstLehre., Abt. 2,* 1953, **31**, 376–380. (*Biol. Abstr.*, **29A**:12995)

SELTZER, C. C. Body disproportions and dominant personality traits. *Psychosom. Med.*, 1946, **8**, 75–97.

SELTZER, C. C. Masculinity and smoking. *Science,* 1959, **130**(3390), 1706–1707.

SHELDON, W. H., DUPERTUIS, C. W., & McDERMOTT, E. *Atlas of men*. New York: Harper, 1954.

SHELDON, W. H., HARTL, E., & MCDERMOTT, E. *Varieties of delinquent youth*. New York: Harper, 1949.

SHELDON, W. H., & STEVENS, S. S. *The varieties of temperament*. New York: Harper, 1942.

SHELDON, W. H., STEVENS, S. S., & TUCKER, W. B. *Varieties of human physique*. New York: Harper, 1940.

SILLS, F., & EVERETT, P. W. The relationship of extreme somatotypes to performance in motor and strength tests. *Res. Quart.*, 1953, **24**, 223–228.

SILLS, F. D. A factor analysis of somatotypes and their relationship to achievements in motor skills. *Res. Quart.*, 1950, **21**, 424–437.

SILLS, R., & MITCHEM, J. Prediction of performance on physical fitness tests by means of somatotype ratings. *Res. Quart.*, 1957, **28**, 64–71.

SIMONSON, E. Outline of analysis of research problems in industrial physiology. *Occup. Med.*, 1947, **3**, 109–134.

SINELNIKOFF, E., & GRIGOROWITSCH, M. The movement of joints as a secondary sex and constitutional characteristic. *Z. KonstLehre., Abt. 2,* 1931, **15**, 679–693.

SMITH, H. C., & BOGARSKY, S. The relationship between physique and simple reaction time. *Charact. Pers.*, 1943, **12**, 46–53.

SPEARMAN, C. *Abilities of man*. London, England: Macmillan, 1927.

STOUDT, H. W., DAMON, A., & MCFARLAND, R. A. Heights and weights of white Americans. *Hum. Biol.*, 1960, **32**, 331–341.

TANNER, J. M. Current advances in the study of physique: Photogrammetric anthropometry and an androgyny scale. *Lancet,* 1951, **240**, 574–579.

TANNER, J. M. The effect of weight and training on physique. *Amer. J. phys. Anthropol.*, 1952, **10**, 427–461.

TANNER, J. M. The reliability of anthroposcopic somatotyping. *Amer. J. phys. Anthropol.*, 1954, 12, 257.

TAPPEN, N. C. An anthropometric and constitutional study of championship weight lifters. *Amer. J. phys. Anthropol.*, 1950, **8**, 49–64.

TAYLOR, W. L., & BEHNKE, A. R. Anthropometric comparison of muscular and obese men. *J. appl. Physiol.*, 1961, **16**, 955–959.

TELKKA, A., PERE, S., & KUNNAS, M. Anthropometric studies on Finnish athletes and wrestlers. *Ann. Acad. Sci. Fenn. Ser. A,* 1951, **28**, 1–34.

THURSTONE, L. L. Current issues in factor analyses. *Psychol. Bull.*, 1940, **37**, 189–236.

THURSTONE, L. L. Analysis of body measurements. *U. Chicago Psychomet. Lab. Rep.*, 1946, No. 29. (a)

THURSTONE, L. L. Factor analyses and body types. *Psychometrika,* 1946, **1**, 15–22. (b)

THURSTONE, L. L. Factorial analysis of body measurements. *Amer. J. phys. Anthropol.*, 1947, **5**, 15–28.

TUCKER, W. B., & LESSA, W. A. Man: A constitutional investigation. *Quart. Rev. Biol.*, 1940, **15**, 265–411.

VIOLA, G. *La constituzione individuale*. Bologna, 1932. 2 vols.

WERTHEIMER, F. I., & HESKETH, F. E. *The significance of the physical constitution in mental disease*. Baltimore, Md.: Williams & Wilkins, 1926.

WIERSMA, E. D. Body build, physiological and psychological functions. *Verh. Akad. Wet. Amst., Ser. B,* 1933, **30**, 1–28.

WIGANT, V. Attempts at anthropometric determination of the body types of Kretschmer. *Acta psychiat. neurol. Scand., Copenhagen,* 1933, **8**, 465–481.

(Received March 12, 1964)

Psychological Bulletin
1964, Vol. 62, No. 6, 427–428

REFLECTIONS OF THE RETIRING EDITOR

HARRY HELSON
Kansas State University

Once every 6 years the editors of APA journals may, upon assuming their office, print statements of policy and/or upon retiring may, as in the *Apologia pro vita sua* of the late Cardinal Newman, seek to justify their ways (toward contributors). There is, of course, nothing to prevent an editor from speaking his piece any time during his incumbency, for editors have freedom to publish or not to publish anything they see fit. Granting that absolute power *may*, but does not have to, corrupt, there is no other way in which men of the desired caliber can be gotten to labor day in and day out, Saturdays, Sundays, holidays, birthdays, anniversaries, and sundry other days, not excepted, as editors of journals. For the job of editing a journal knows neither length nor kind of day. This power, especially in the case of journals which are considered to be "quality," where merely to publish an article impresses Deans and Presidents, is one to be exercised with the utmost objectivity and restraint. The power to deny publication of whole articles or parts thereof is literally the power of life or death not only to individuals but also to facts, points of view, theories, movements, topics, areas, and ideas. No wonder editors and their anonymous consultants are probably the most feared psychologists in the profession. They are not loved for the good they do because nobody (hardly anyone, that is) feels grateful to an editor for accepting an article as it stands, or for suggestions toward improving it, rewriting it, or for giving various and sundry bits of advice, such as deleting *ad hominem* phrases. Positive feelings towards editors are completely hidden or expressed at most in private communications, while negative feelings may even break into print, if authors have not achieved catharsis in personal correspondence.

Every contributor, with few exceptions, tends to think of himself and his article as unique. He is intensely interested in his subject, or at least in getting something published about it; he has spent a lot of time and effort in reading and writing (usually); he has gone to some expense in having it typed (if it is not supported by a contract, grant, or the department budget); and he has been working on it for months, perhaps years, and wants to see it in print at the earliest possible moment after it reaches the editor's desk. When asked to reduce its length because the article is repetitious, padded, disjointed, amorphous, or because the publication lag in the journal is unconscionably long, he feels put upon or that he is the victim of circumstances that could easily be remedied if the editor were not so persnickety, the Association more aware of his rights as a member, or if so-and-so's article which appeared in such-and-such an issue had been subjected to the same strictures as his. The individual contributor does not know how his presentation impresses others because this is something that even his best friend will not tell him. The individual contributor is not expected to have and does not have any feeling for balancing areas, topics, and treatments to provide a fair sampling of what is going on in the vast enterprise called "Psychology." It remains for editors to perform this task in most cases. To the contributor there is never a question of the quality or inherent worth of his topic and treatment, for no matter how

trifling, tenuous, or weak the publications in an area may be, the one who is immersed in that area finds it solid, substantive, and satisfying, or he would not take the time and effort to write about it.

I have mentioned balance and fair sampling as considerations in publishing a journal that is open to all areas of psychology. There are others which are even more important both to readers of the journal and the progress of psychology which have little or no appeal when they affect the individual contributor adversely. *Sometimes,* rarely to be sure, but occasionally, an article is published that seems to violate the editorial policy of the journal, and this is because some larger consideration is involved. Among such considerations are the possibility of inciting others to work in a new area, originality or freshness of treatment, stand-out quality, or paucity of material in the area represented by the article. The reverse side of the coin concerns rejection of articles because of a plethora of contributions in a given area, pedestrian treatment, or unsuitability in a learned journal. In the latter connection it may be pointed out that what is perfectly good and proper for a chapter in a textbook or for layman's consumption in a seminar may be wholly out of place in a learned journal cramped for space. Suitability for publications, balance of areas and topics, and amount of space available concern matters that only one having an overall view can properly assess, and here the advantage lies with the editor rather than with the writer of a single article.

There is, however, one area in which the contributor's decision should take precedence over the editor's and his consultants, and that is the interpretation, theory, or hypothesis used to organize or make meaning out of disparate facts. It is proper to ask an author to consider other interpretations and to point out weaknesses in his point of view, but to deny him the privilege of presenting his theory in print is to suppress ideas which, baldly, is a form of book burning. Much as this editor has been opposed to certain approaches, points of view, theories, and interpretations, he has never been responsible for their deletion from any contribution. Criticisms of authors' interpretations have been passed on to them and, if the article was otherwise acceptable, it was left to them whether or not they changed their interpretations. After all, authors have rights and the right to explain his data is one of the most sacred the scientist possesses. Let no editor infringe it. Advances in science come only through new theories, new interpretations, new ideas.

Only in the case of *ad hominem* arguments has this editor insisted that contributors modify their remarks. Criticisms, rebuttals, and counterrebuttals should be addressed to facts, ideas, and problems, not to persons. Nor does snide innuendo have any place in scientific writing.

On the whole, editing *Bulletin* for 6 years has been a rich and rewarding experience. Apart from having been accused a *few* times of this or that by members of APA (not always contributors) my relations with contributors, Central Office personnel, and others have been friendly, often cordial. I shall miss the contacts being editor has afforded me but I shall not miss the mountain of manuscripts, letters, and sundry materials that used to greet me after an absence of more than a few days from my duties. So "Hail and Farewell" ye authors, whether your contributions have been accepted or rejected, and may you all remember me kindly and with sympathy, if you remember me at all!